

Scientific Computing with Amazon Web Services

Jamie Kinney

Sr. Manager of Scientific Computing Amazon Web Services

jkinney@amazon.com

@jamiekinney

<http://bit.ly/AWSOSG>



Amazon Global Impact Initiatives

Scientific Computing

- Global “Big Science” Projects
- Enabling the “long tail of science”
- Collaborative research
- Accelerating the transition to “Networked Science”

Open Data

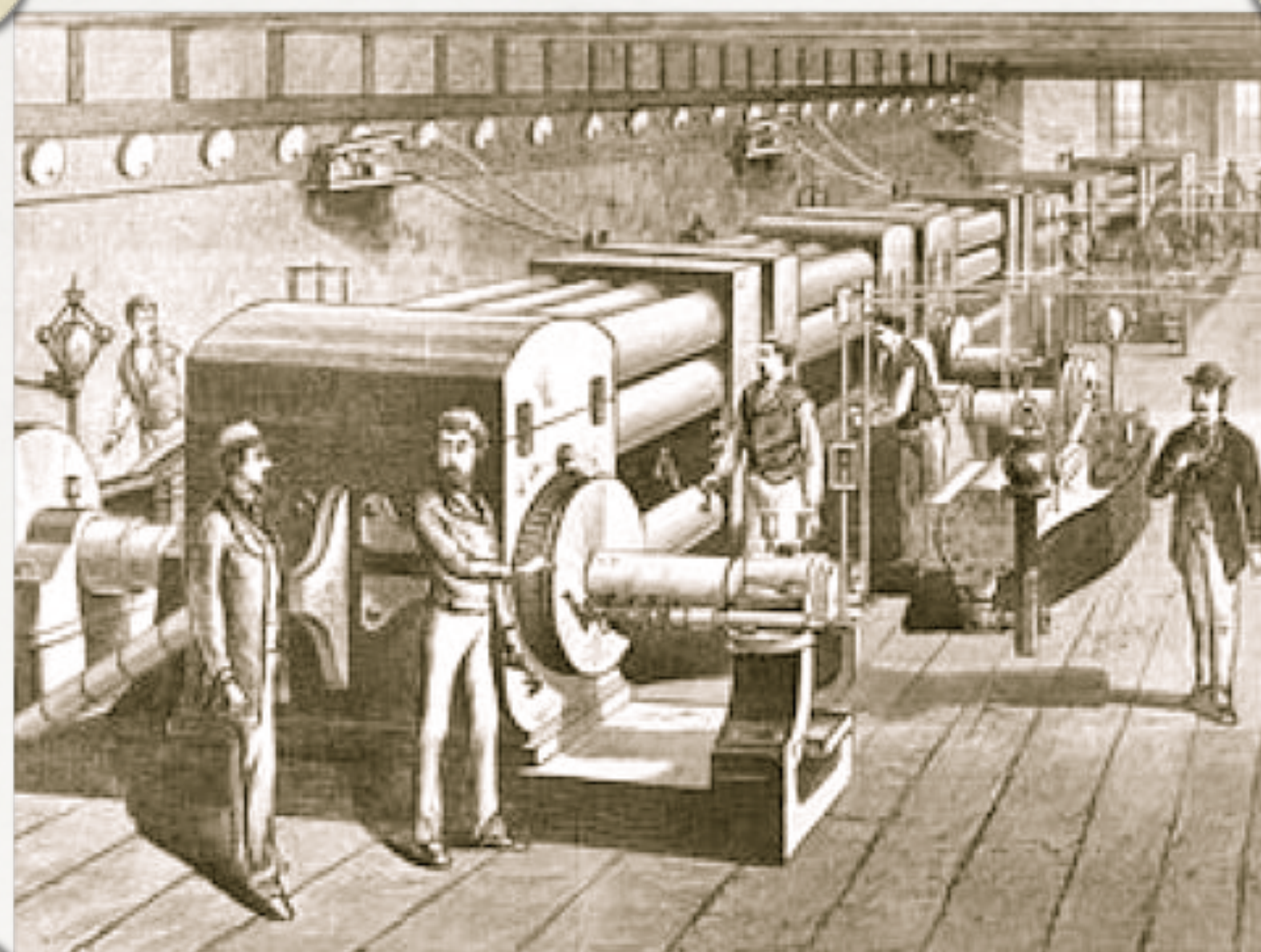
- AWS as the platform for open-access data portals and archives
- Amazon Public Data Sets
- Public/Private Data partnerships

Economic Development

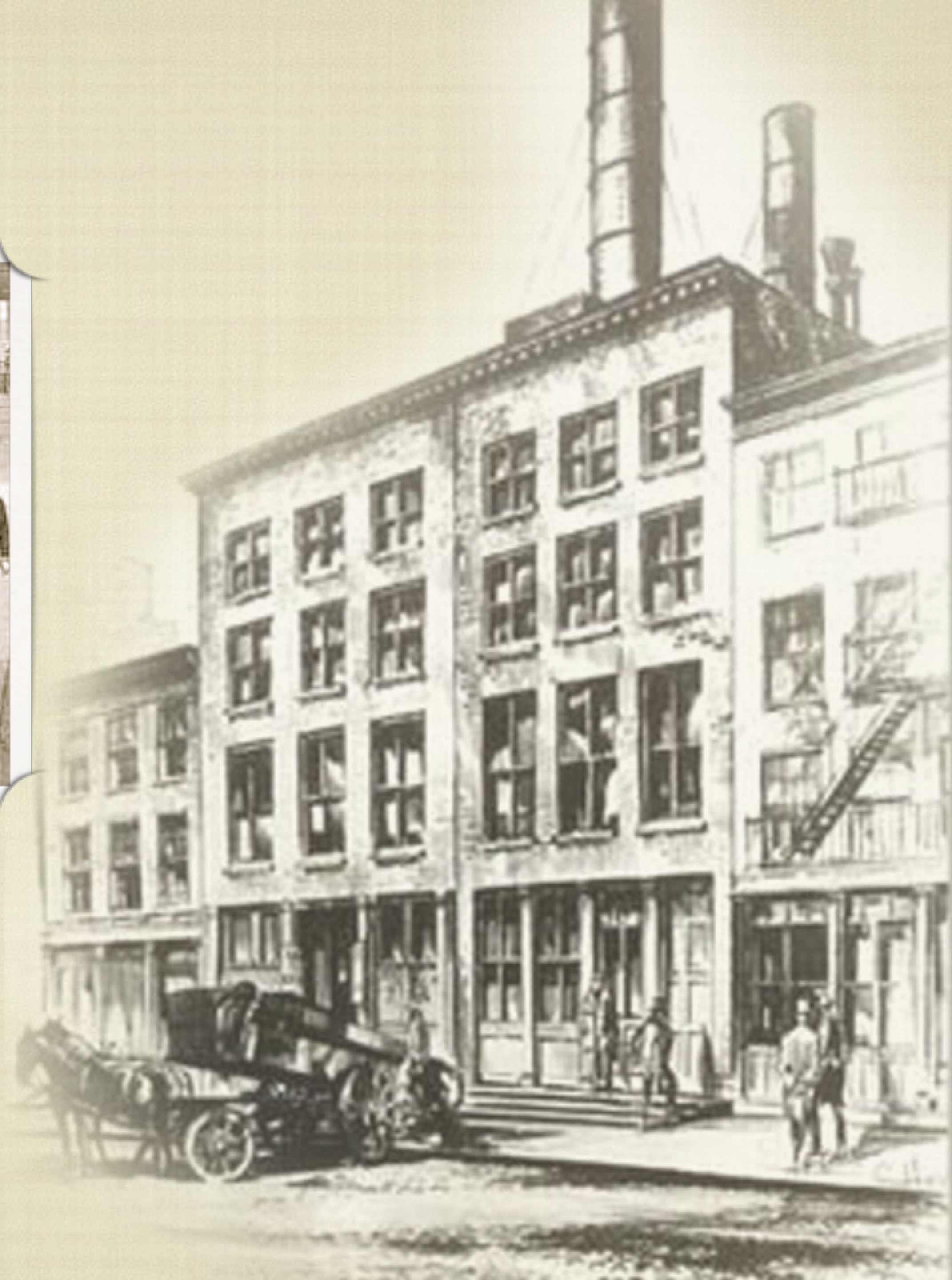
- The Amazon Job Accelerators program

Why are we focusing on the Scientific Community?

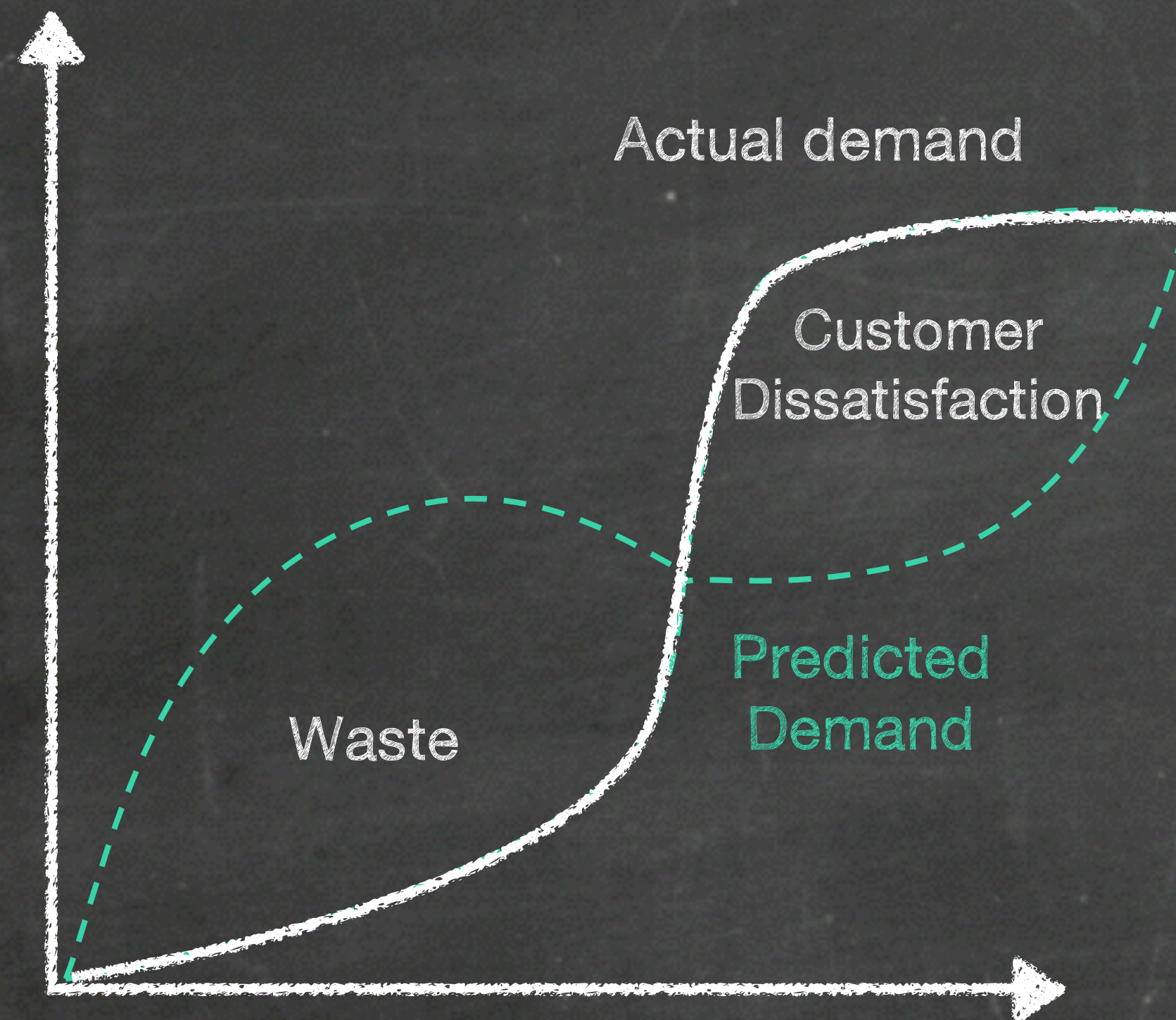
- Scientific computing is a profitable business for AWS
- To accelerate the pace of scientific discovery
- To develop new capabilities which will benefit all AWS customers
 - Streaming data processing & analytics
 - Exabyte scale data management solutions
 - Collaborative research tools and techniques
 - New AWS regions (e.g. South Africa and Western Australia for the SKA)
 - Significant advances in low-power compute, storage and data centers
 - Identify efficiencies which will lower our costs and pricing for customers
 - Push our existing services to support exabyte/exaflop scale workloads



Pearl Street
Power Station



Self Hosting



Rigid



Elastic

2003

amazon.com

\$5.2B retail business
7,800 employees
A whole lot of servers

2013



Every day, AWS adds enough
server capacity to power this
\$5B enterprise

Regions

Availability Zones

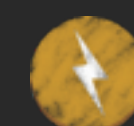
Content Delivery POPs



Networking



VPC



Direct Connect



Route 53

Compute



EC2



Elastic Load Balancer



Auto Scaling



S3



EBS



Glacier



Storage Gateway



Import/Export

Storage



RDS

MySQL, PostgreSQL
Oracle, SQL Server



DynamoDB



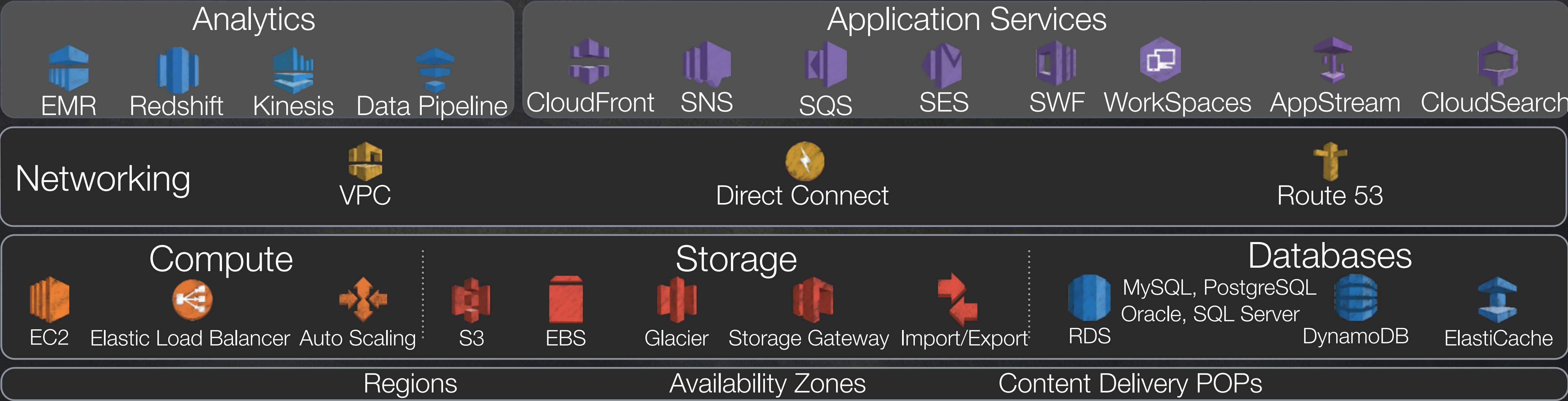
ElastiCache

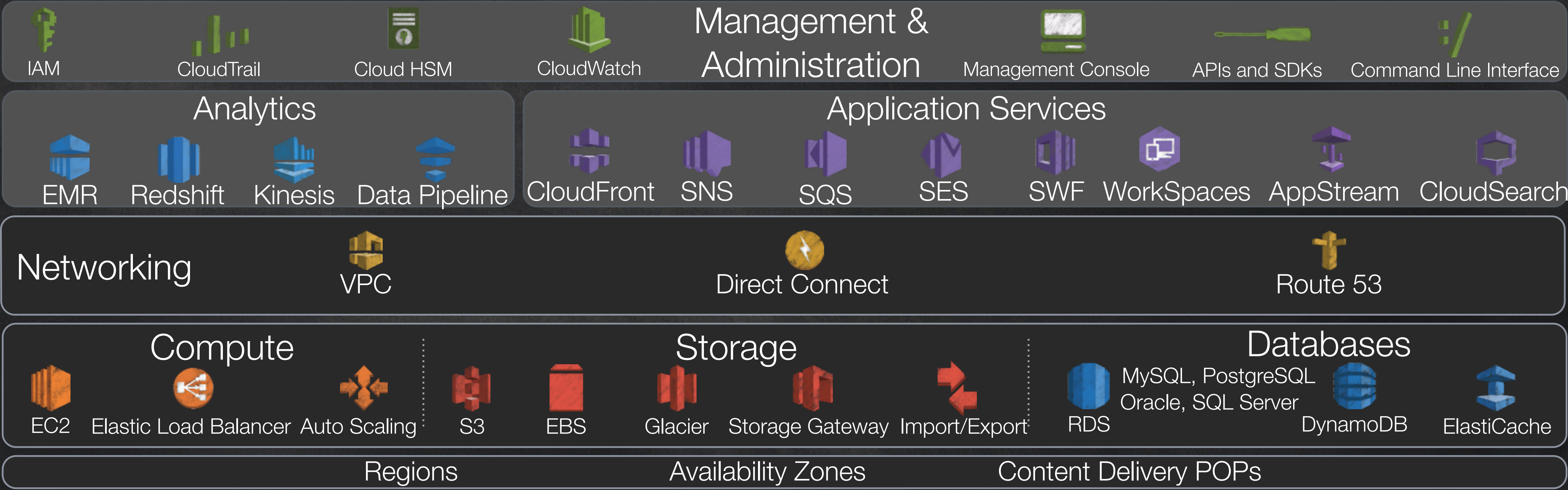
Databases

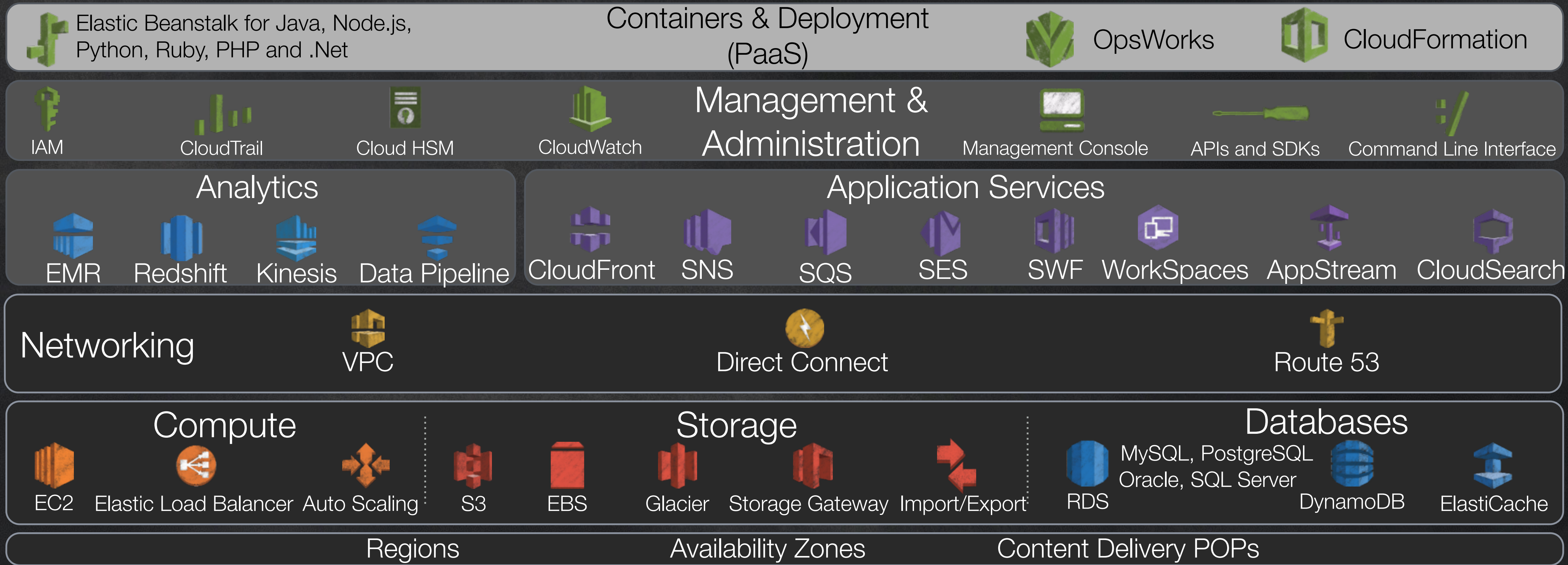
Regions

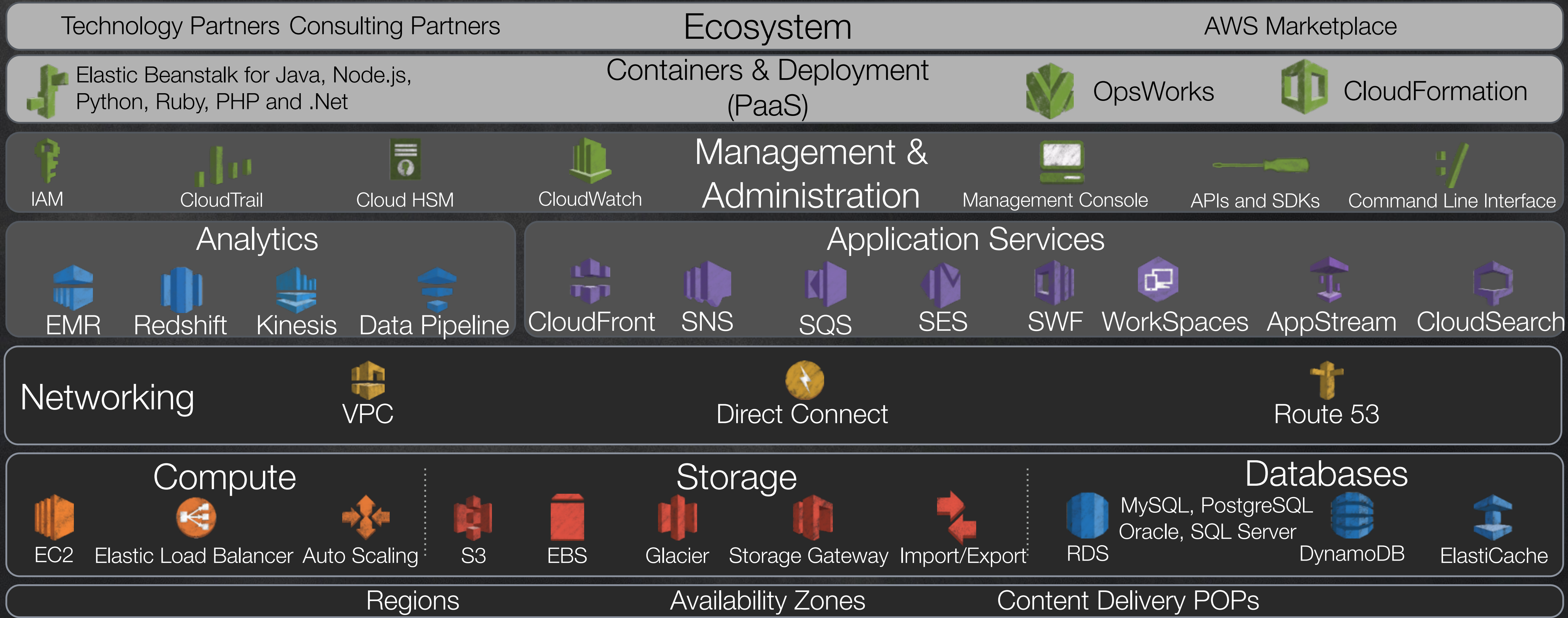
Availability Zones

Content Delivery POPs









Support

Professional Services


Training

Certification

Technology Partners Consulting Partners

Ecosystem

AWS Marketplace



Elastic Beanstalk for Java, Node.js, Python, Ruby, PHP and .Net

Containers & Deployment (PaaS)



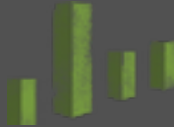
OpsWorks



CloudFormation



IAM



CloudTrail



Cloud HSM



CloudWatch

Management & Administration



Management Console



APIs and SDKs



Command Line Interface

Analytics



EMR



Redshift



Kinesis



Data Pipeline

Application Services



CloudFront



SNS



SQS



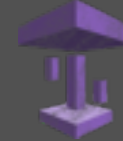
SES



SWF



WorkSpaces



AppStream



CloudSearch

Networking



VPC

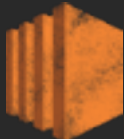


Direct Connect



Route 53

Compute



EC2



Elastic Load Balancer



Auto Scaling



S3



EBS



Glacier



Storage Gateway



Import/Export

Storage



RDS

MySQL, PostgreSQL
Oracle, SQL Server



DynamoDB



ElastiCache

Databases

Regions

Availability Zones

Content Delivery POPs

Broad Features and Functionality for Each Service



EC2



Elastic Load Balancer



Data Pipeline



IAM



AppStream



EBS



CloudFront



Route 53



Redshift



S3



Elastic Map Reduce



CloudFormation



DynamoDB



Kinesis

56 new features
since Feb 2013

Regional expansion to US West (Oregon)
Support for temporary credentials when loading data from Amazon S3
Regional expansion to EU West (Dublin)
SOC1/2/3 Compliance certification
Ability to UNLOAD encrypted files in parallel to Amazon S3
Regional expansion to Asia Pacific (Tokyo)
Support for JDBC fetch size to enable extraction of large data sets over JDBC/ODBC
Enable logging of UNLOAD statements
New built-in function to compute the SHA1 hash of a value
Added support for UTF-8 characters up to 4 bytes in size
Ability to share snapshots between accounts to simplify manageability.
Support for statement timeouts to automatically terminate queries that exceeded allotted execution time
Added support for timezone conversion in SQL
Added support for datetime values expressed in milliseconds since EPOCH to simplify ingestion
Simplified ingestion by automatically detecting date and time formats.
Added support for automatic query timeouts to workload management queues.
Enabled the use of wildcards when assigning queries to workload management queues.
New built-in function to enable customers to calculate the CRC32 checksum of a value
Console improvements to show progress bars for backup and restore operations.
Added the ability to support IAM at the resource level allowing tight control of who can take what actions on which resources.
Obtained PCI compliance
Added the ability to substitute a customer chosen character for invalid UTF-8 characters to simplify ingestion
Allowed customers to store JSON data in VARCHAR columns and added built-in functions to enable data extraction
Added support for POSIX regex expressions when using SIMILAR to in SQL queries
Added Cursor support to enable extraction of large data sets over ODBC connections
Built-in function to enable splitting a string using a supplied delimiter to make parsing values easier
Added system tables to enable logging of database activity for auditing
Regional expansion to Asia Pacific (Singapore, Sydney)
Enable customers to control cluster encryption keys by using an on premises hardware security module (HSM) or Amazon CloudHSM
Enable customers to receive alerts via SNS for informational or error-related events for cluster monitoring, management, configuration and security.
Integration with Canal to enable streaming data ingestion
Copy from an arbitrary SSH connection enabling direct copy from Amazon EMR, HDFS, or any other database that supports SSH access and script execution
Enable distributing tables to all compute nodes to speed up queries, especially those involving star or snowflake schemas
Logging of database logins, failed logins, SQL execution and data loads to S3 and integration with CloudTrails for control plane events
Enabled caching of database blocks to speed up access to frequently queried data
Increase cluster concurrency limits from 15 to 50 to enable higher concurrent query execution
Optimizations to resize code that lead to 2-4x improvement in resize performance
Approximate COUNT DISTINCT using HyperLogLog giving 10-20x performance improvements with less than 1% error
Enable customers to continuously, automatically and incrementally back up data to a second AWS region for DR
On track to obtain Fedramp certification
Deliver Redshift on SSD instances enabling a lower-cost, high performance entry point

AWS Big Data Technologies

Amazon RDS



Hosted
Relational
Databases

Amazon DynamoDB



Managed
NoSQL
Database

Amazon ElastiCache



In-memory
Caching Service

Amazon Elastic Map Reduce



Hosted Hadoop
framework

Amazon Data Pipeline



Move data
among AWS
services and on-
premise data
sources

Amazon Redshift



Petabyte-scale
columnar
relational data
warehouse
service

<http://aws.amazon.com/big-data/>

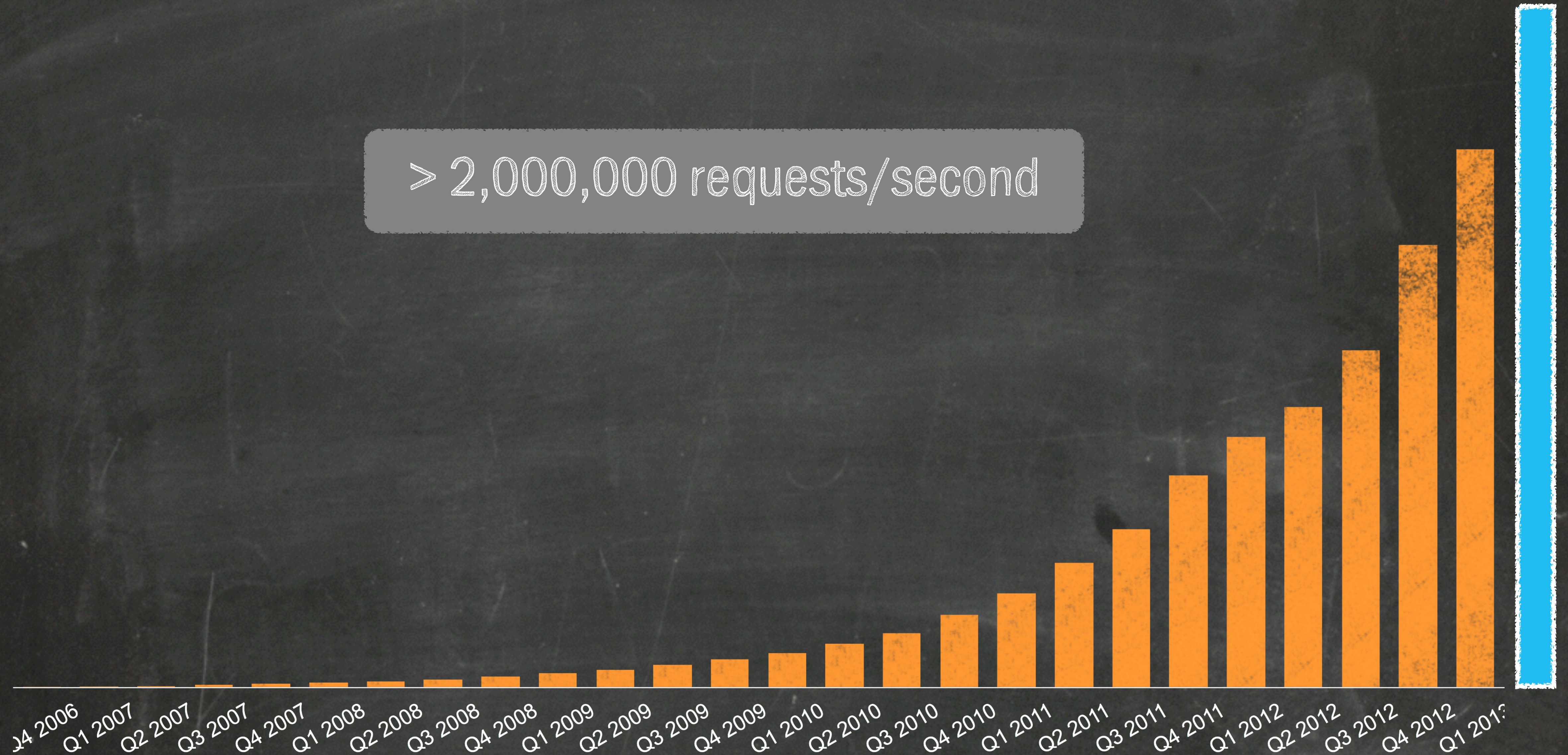
The AWS Big Data Stack Also Includes...

- HTCondor
- Spark/Shark
- Cassandra
- MongoDB
- Hadoop
- Pegasus
- SLURM
- MIT StarCluster
- SNAP
- BlinkDB
- Galaxy
- Lustre
- Gluster
- OrangeFS
- BOINC
- ... and many more

Amazon S3

Amazon S3: Over 2 Trillion Total Objects

> 2,000,000 requests/second



Amazon EC2

EC2 HPC Instance Types

instance type	features	use case
cg1 - GPGPU	2 NVIDIA M2050 2.93GHz Nehalem 22.5gb RAM 2x840gb local storage	GPU based computing CUDA & OpenCL rendering engineering design
hi1 - high IO / SSD	120k random IOPS Intel Xeon 60.5gb RAM 2 x 1tb SSD	databases shared filesystems high IOPS computing
hs1 - storage	48 TB raw storage Intel Sandy Bridge 117gb RAM 24 x 2tb local storage	large scale data storage node cluster filesystem data warehousing
cr1 - memory	244gb RAM, AVX, AES-NI 2.6GHz Intel Sandy Bridge with Turbo 2 x 120gb SSD	in memory analytics large cache large memory hoc genome assembly and analytics

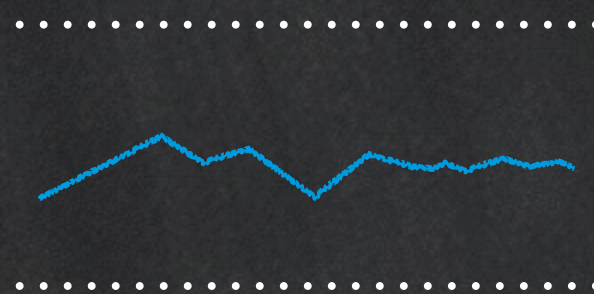
EC2 HPC Instance Types - continued

instance type	features	use case
cc2 - compute	Intel Turbo AVX and AES-NI 2.6GHz Sandy Bridge 60.5 gb RAM 4x840gb local storage	general purpose computing HPC CFD/CAE/MM
g2 - graphics	NVIDIA GK-104 2.6GHz Sandy Bridge 15 gb RAM 60 gb SSD	remote visualization HPC pre-post processing 3d rendering
hi2 - iops / SSD	300,000 random IOPS 2.6 GHz Intel Ivy Bridge - 2/4/8/16/32 VCPU up to 244 gb RAM up to 5.8 tb SSD	large scale data storage node cluster filesystem data warehousing
c3 - compute	turbo to 3.4 Ghz 2.5GHz Ivy Bridge - 2/4/8/16/32 VCPU up to 64 gb RAM up to 640 gb SSD	general purpose computing HPC CAE/CFD/MM
r3 - memory	turbo to 3.4 Ghz 2.5GHz Ivy Bridge - 2/4/8/16/32 VCPU up to 244 gb RAM up to 640 gb SSD	general purpose computing HPC Memory-intensive computing

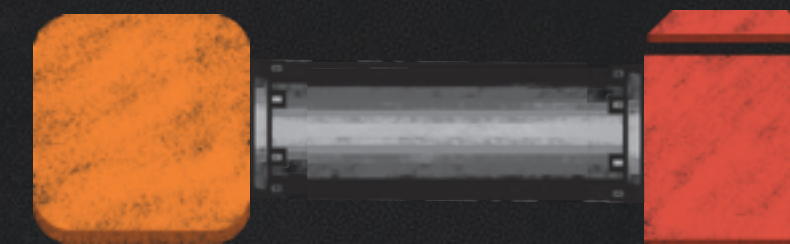
...Connected by a World-class Network



High packets-per-second performance



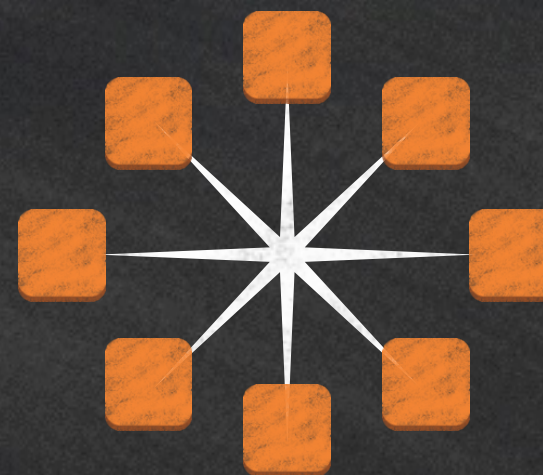
Low jitter



EBS-optimized instances



Virtual network interfaces



High throughput,
low latency



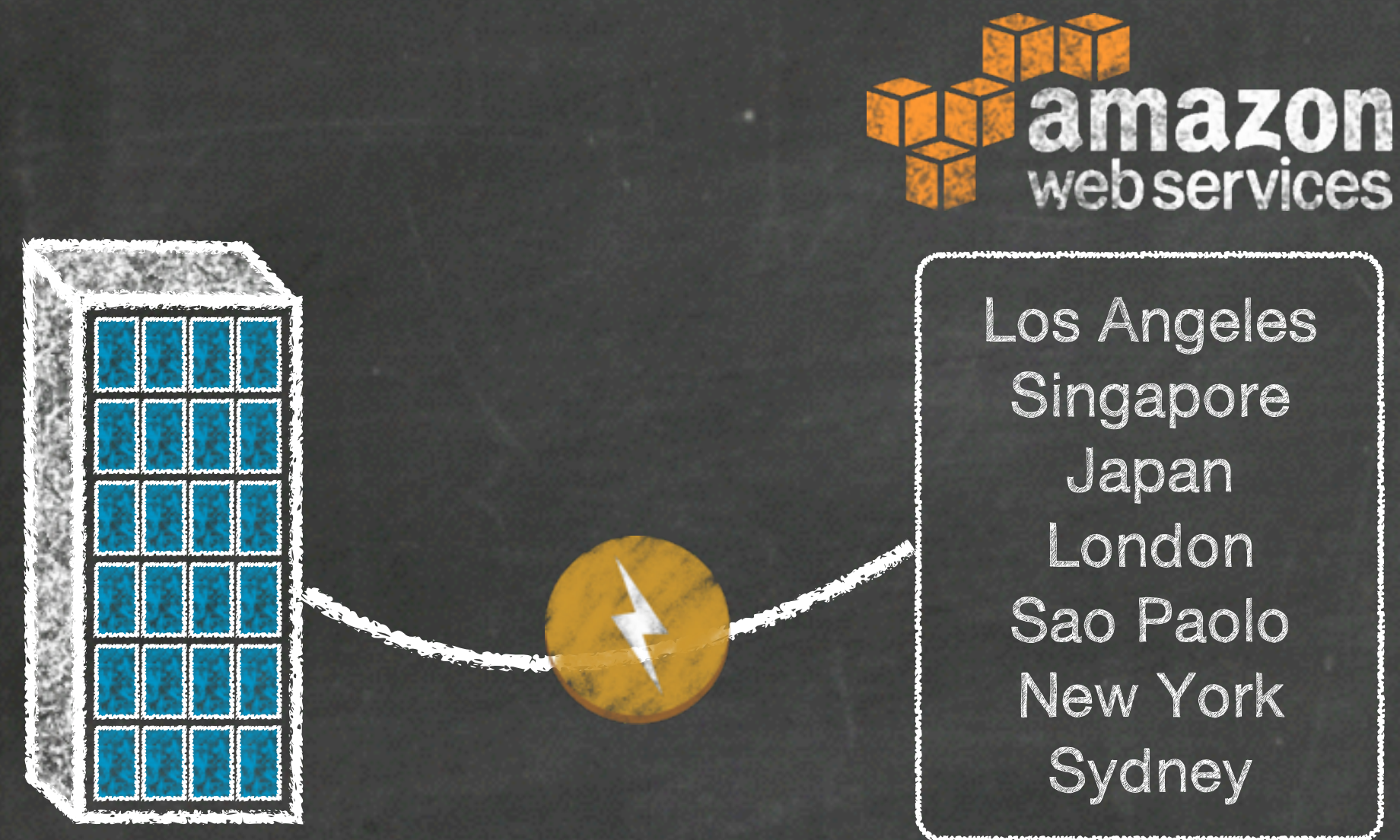
Physical placement optimization

#64 on Top500

	Academy of Sciences China	GHz, Infiniband QDR, NVIDIA 2050 IPE, Nvidia, Tyan				
64	Amazon Web Services United States	Amazon EC2 C3 Instance cluster - Amazon EC2 Cluster, Intel Xeon E5-2680v2 10C 2.800GHz, 10G Ethernet Self-made	26496	484.2	593.5	
65	United Kingdom Meteorological Office United Kingdom	Power 775, POWER7 8C 3.836GHz, Custom Interconnect IBM	18432	476.3	565.6	1040

<http://top500.org/list/2013/11/>

Integrated Architectures

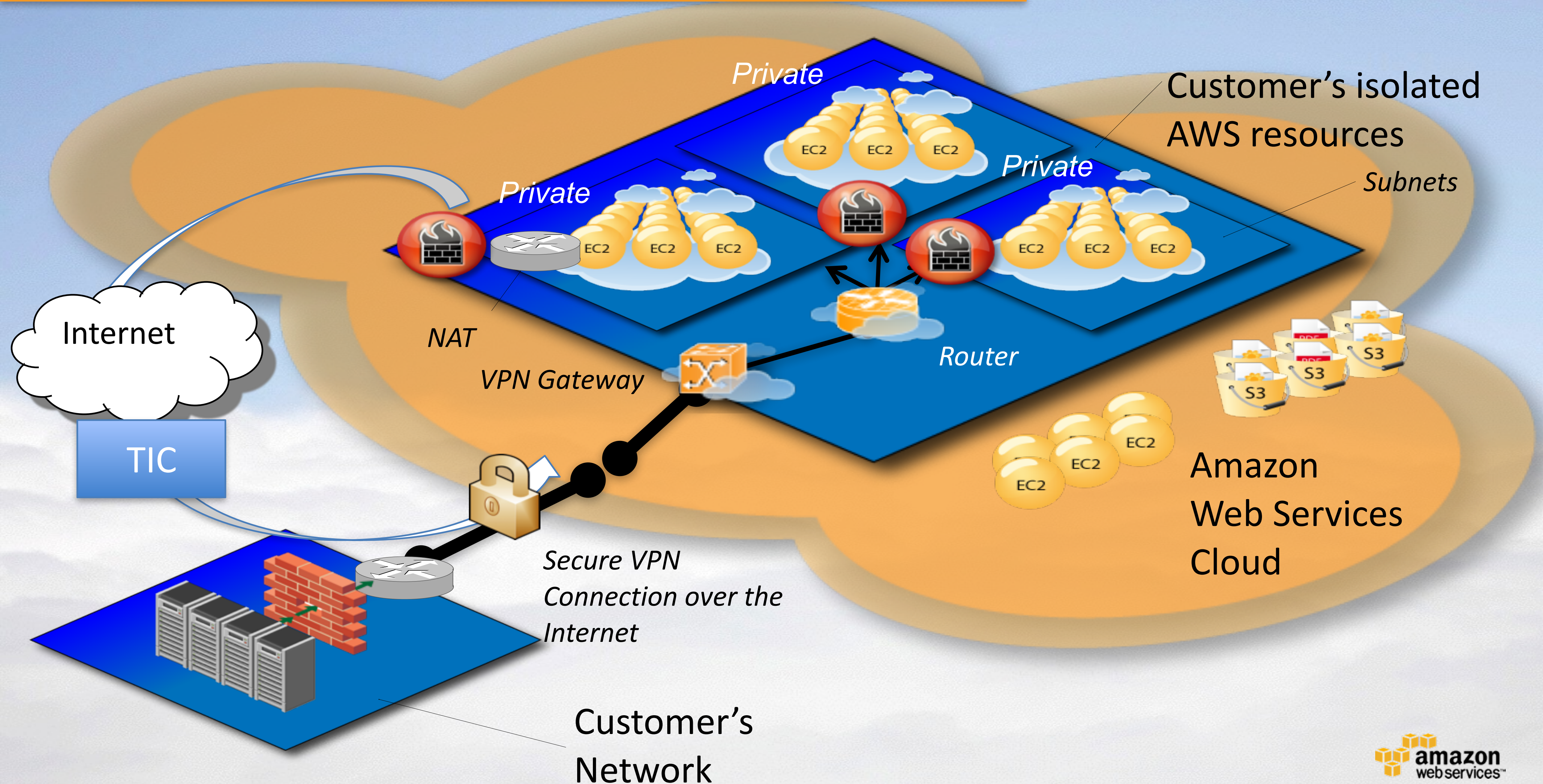


**AWS Direct
Connect**



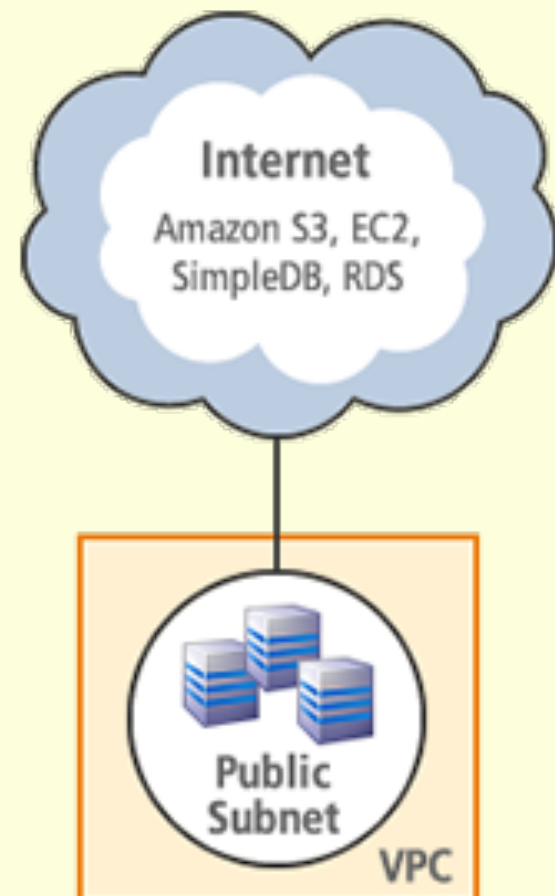
Amazon VPC

AWS Virtual Private Cloud Networking



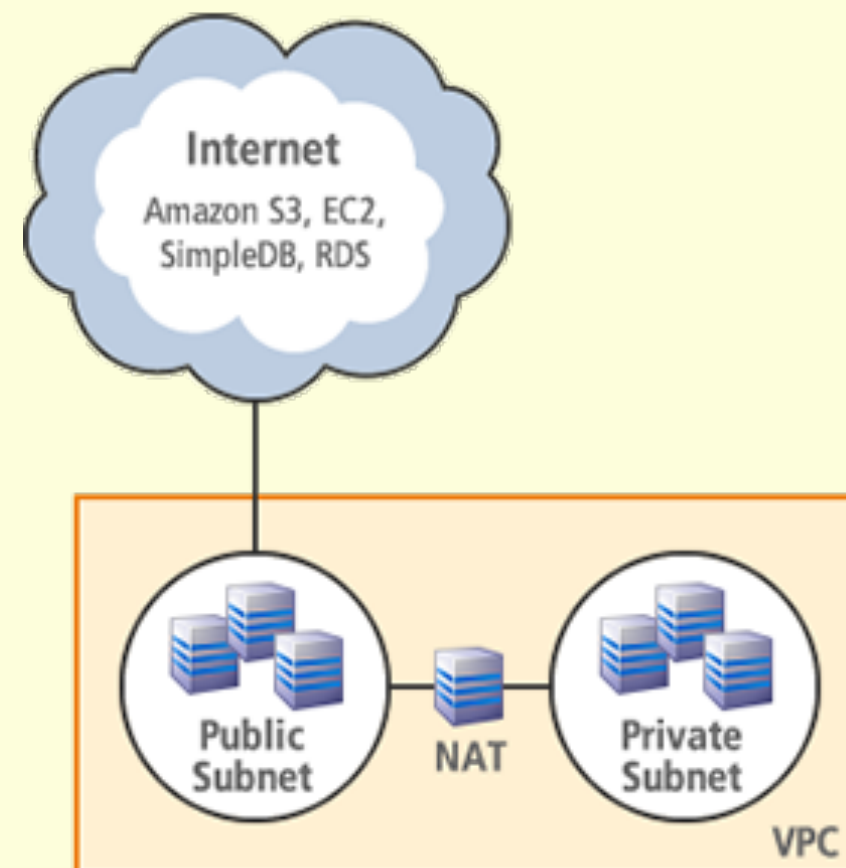
VPC Deployment Models

VPC With a Single Subnet



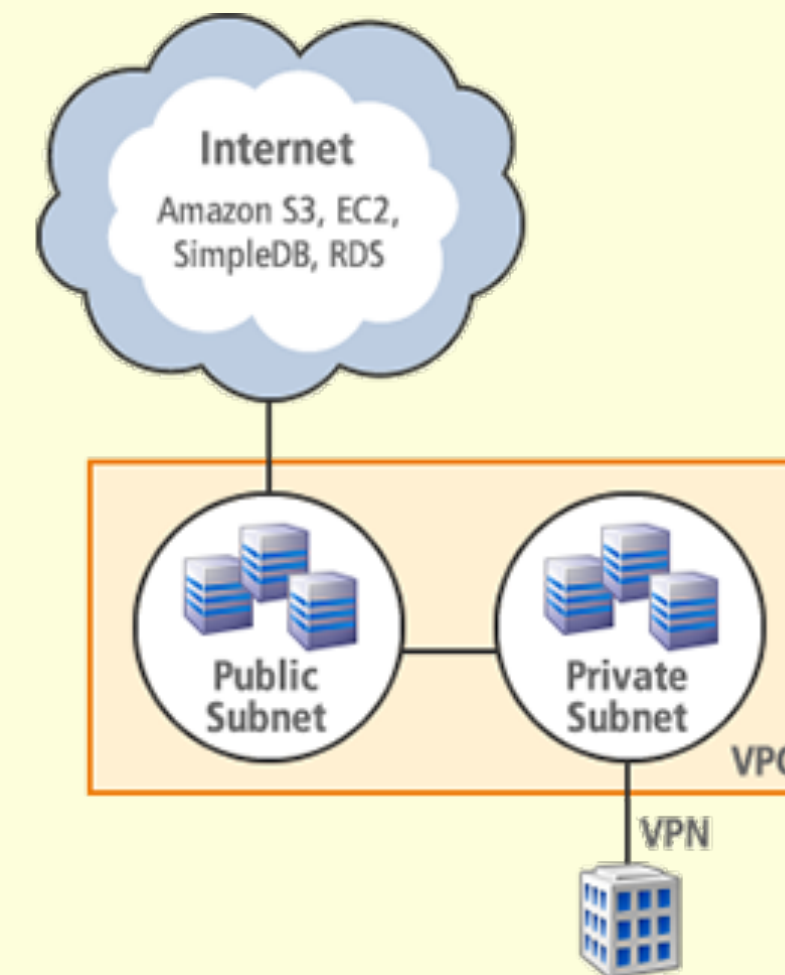
Simple web hosting, collaborative research environments, on-demand HPC/HTC clusters

VPC With Private and Public Subnets



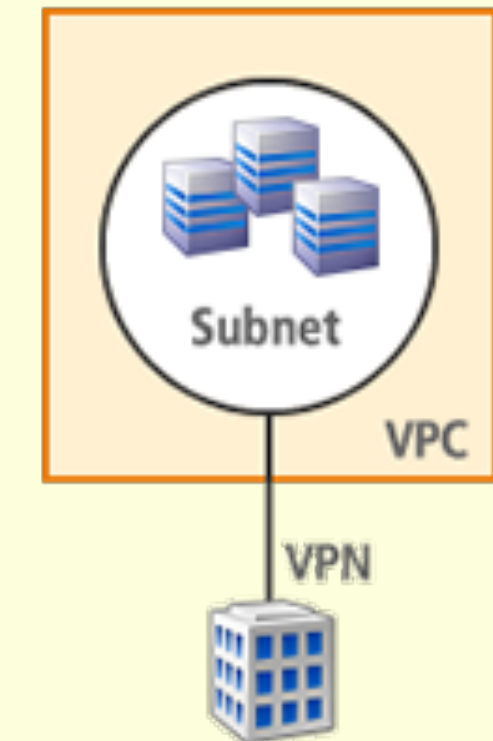
Multi-tier web hosting

VPC With Private and Public Subnets & Hardware VPN access





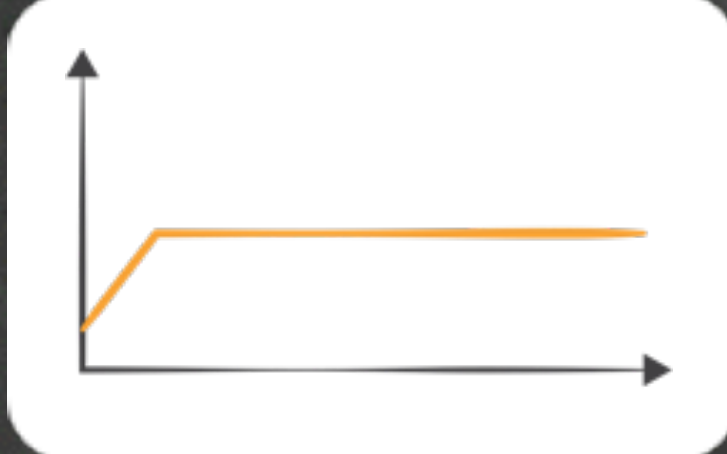
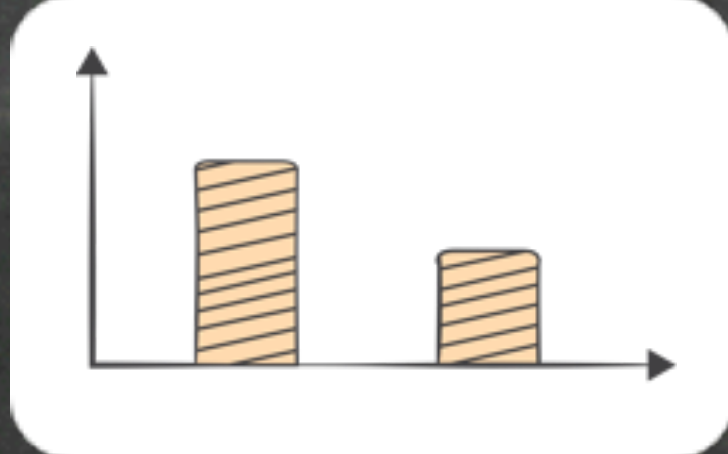
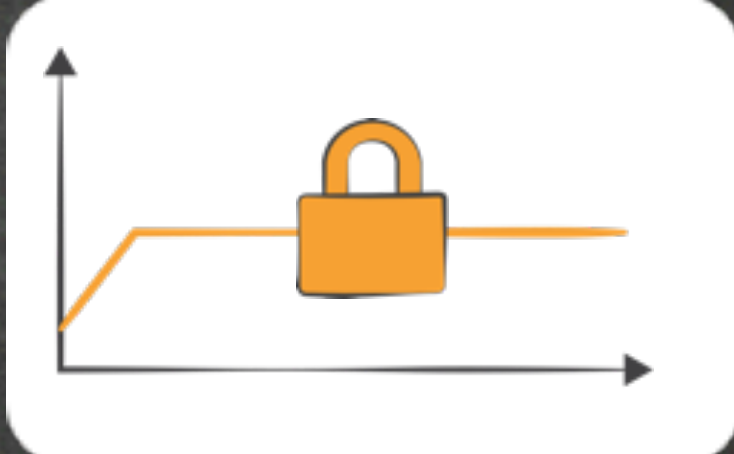
Multi-tier web hosting with access to internal infrastructure

VPC With a Private Subnet Only & Hardware VPN Access



Seamless private expansion of on-premise infrastructure. "Burst capacity" or dedicated cloud environments with connectivity to on-premise resources

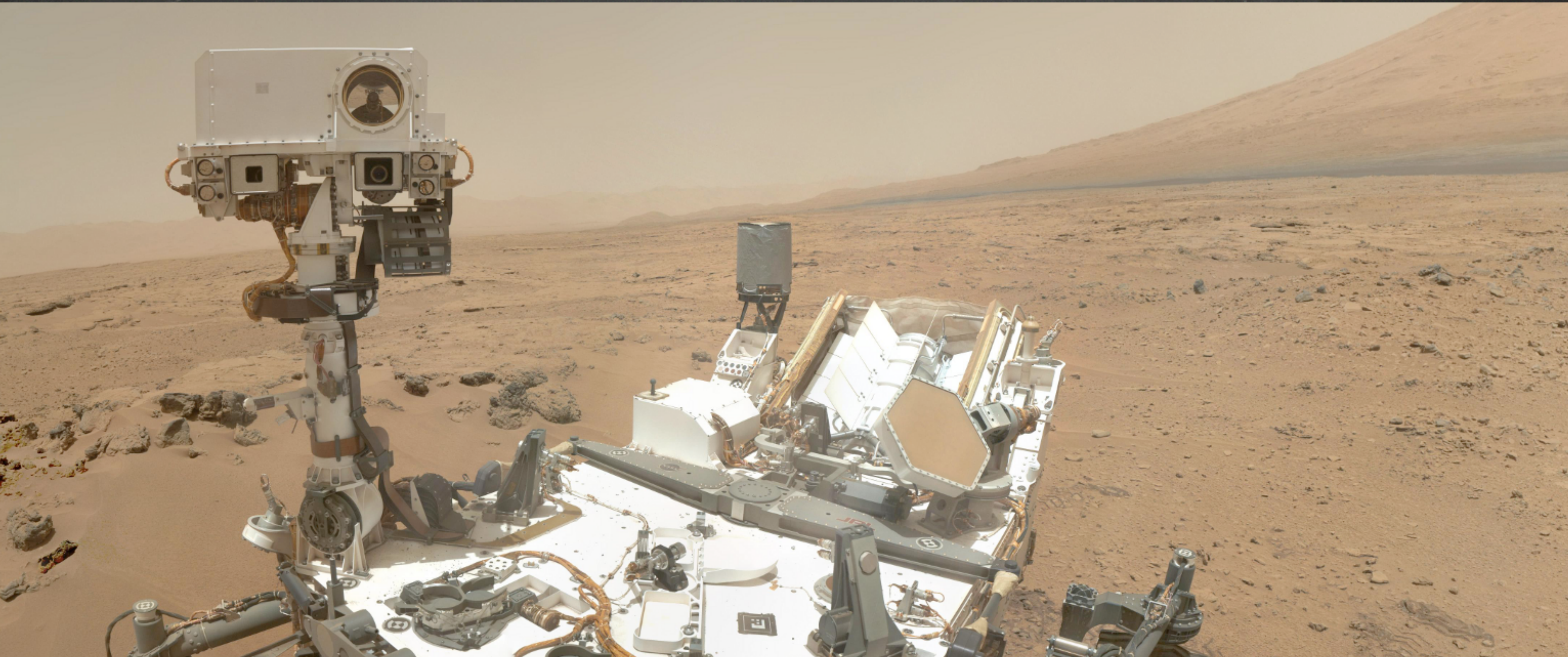
Multiple Purchase Models

free tier	on-demand	reserved	spot	dedicated
get started on AWS with free usage & no commitment	pay for compute capacity by the hour with no long-term commitments	make a low, one-time payment and receive a significant discount on the hourly charge	bid for unused capacity, charged at a Spot Price which fluctuates based on supply and demand	launch instances within Amazon VPC that run on hardware dedicated to a single customer
for POCs and getting started	for spiky workloads, or to define needs	for committed utilization	for time-insensitive or transient workloads	for highly-sensitive or compliance-relayed workloads
				

Scientific Computing Use Cases

- Science-as-a-Service
- Large-scale HTC (100,000+ core clusters)
- Large-scale MapReduce (Hadoop/Spark/Shark) using EC2 or EMR
- Small to medium-scale clusters (hundreds of nodes) for traditional MPI workloads
- Many small MPI clusters working in parallel to explore parameter space
- Small to medium scale GPGPU workloads
- Dev/test of MPI workloads prior to submitting to supercomputing centers
- Ephemeral clusters, custom tailored to the task at hand, created for various stages of a pipeline
- Collaborative research environments
- On-demand academic training/lab environments

Who is using AWS?



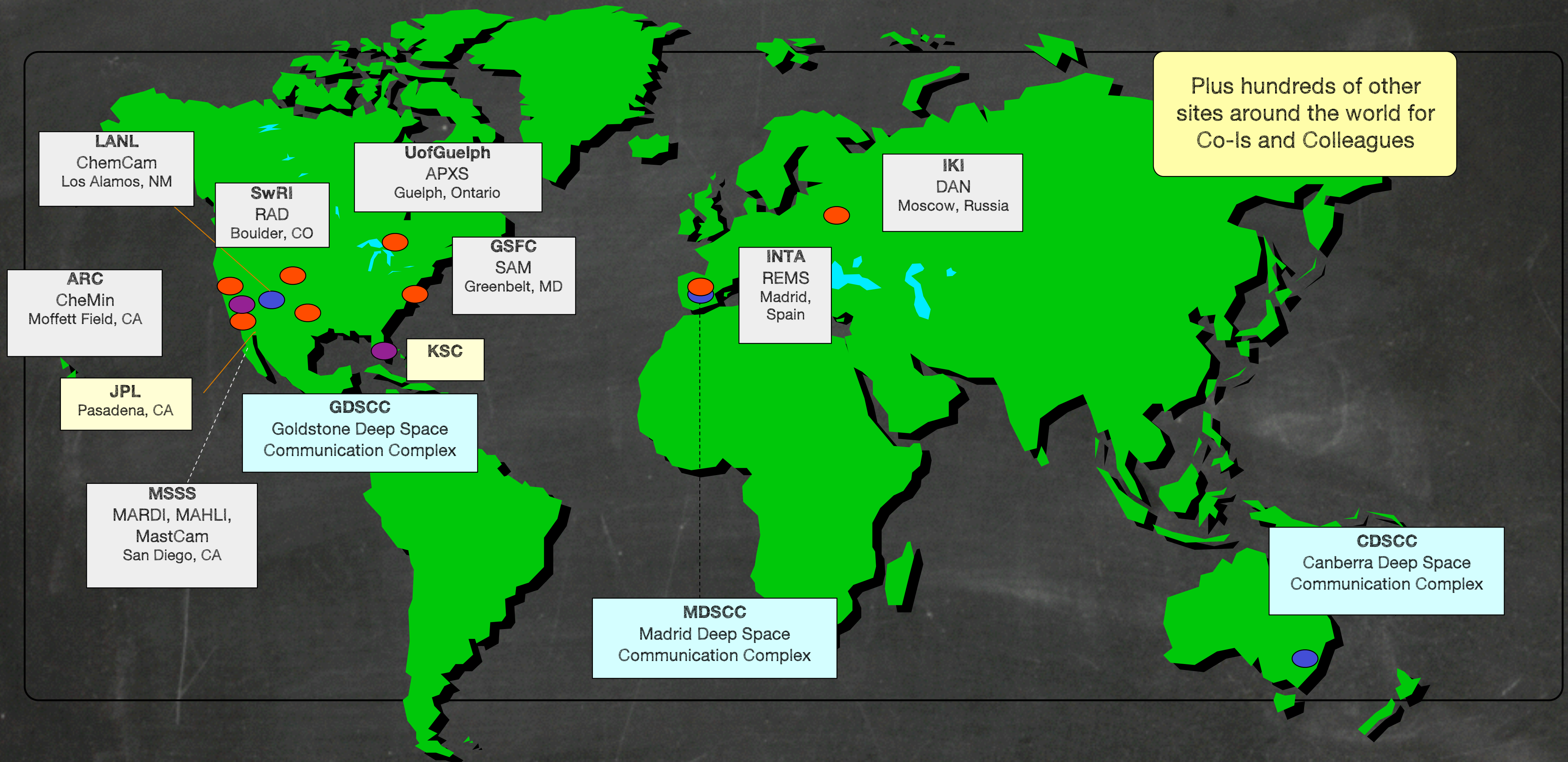


Map of scientific collaborations from 2005 to 2009

Computed by Olivier H. Beauchesne @ Science-Metrix, Inc.

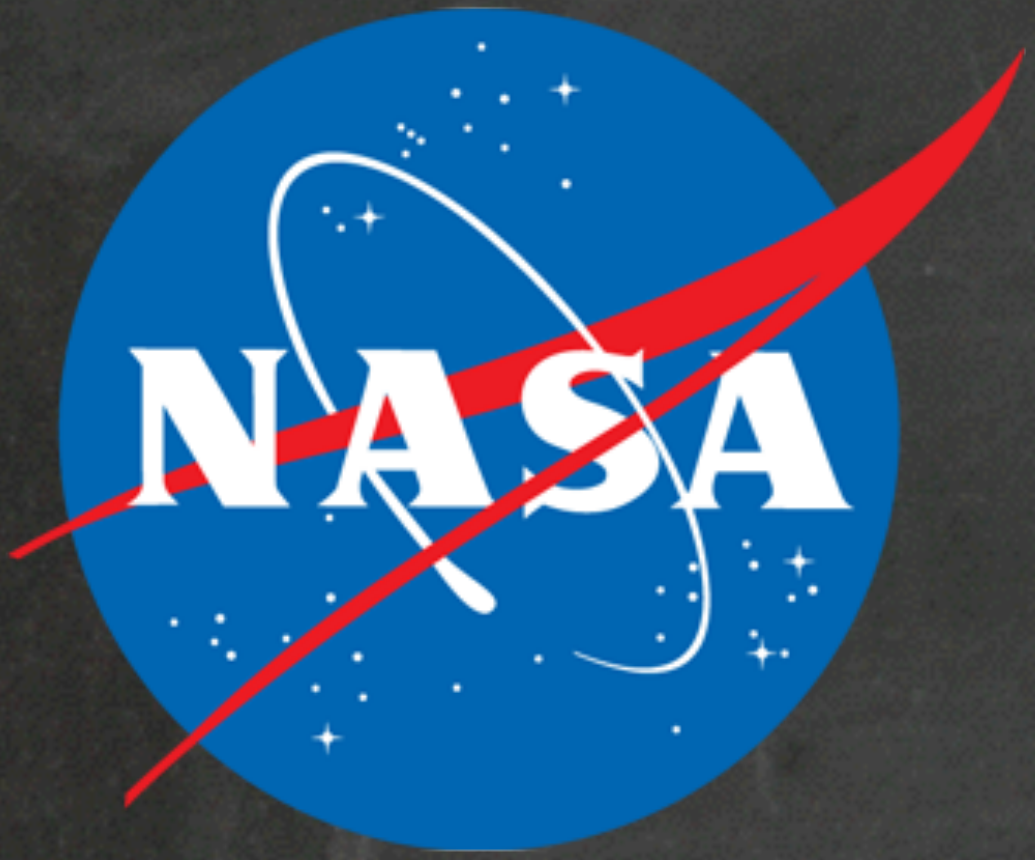
Data from Scopus, using books, trade journals and peer-reviewed journals

MSL Distributed Operations



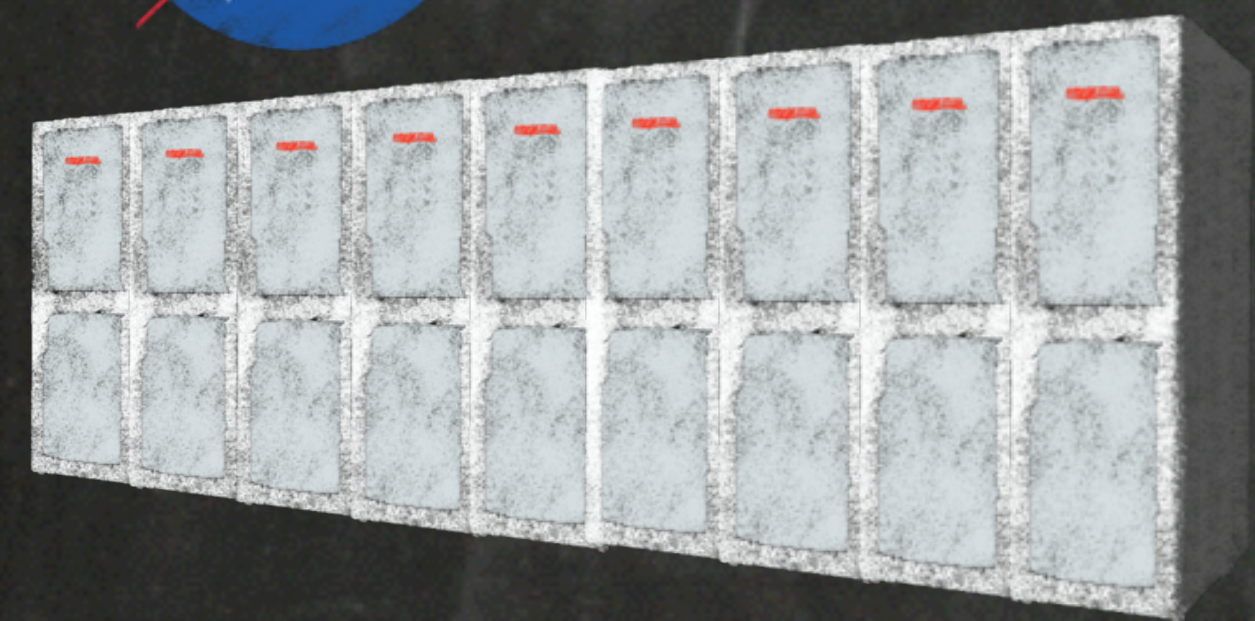
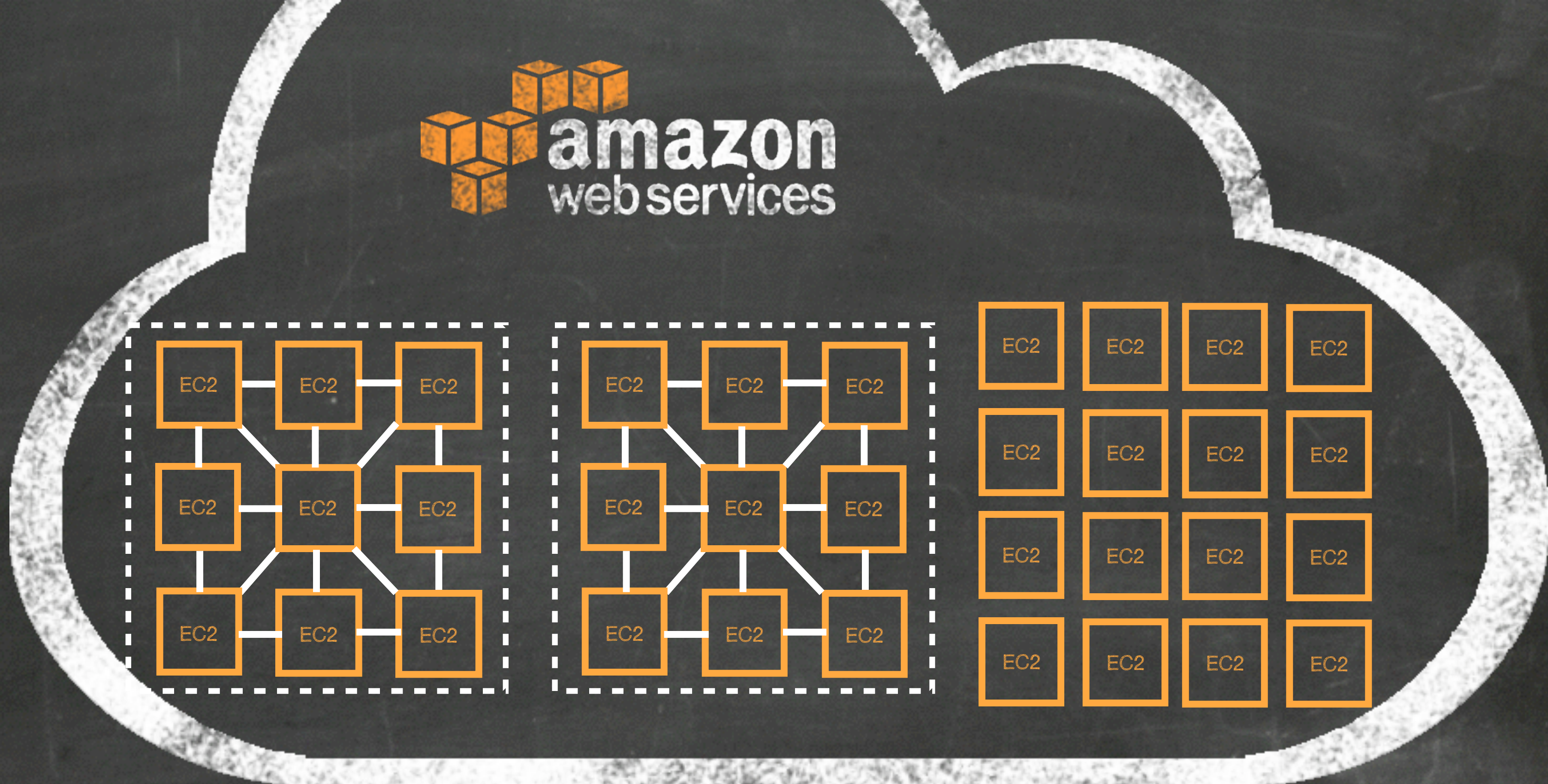






Jet Propulsion Laboratory

California Institute of Technology

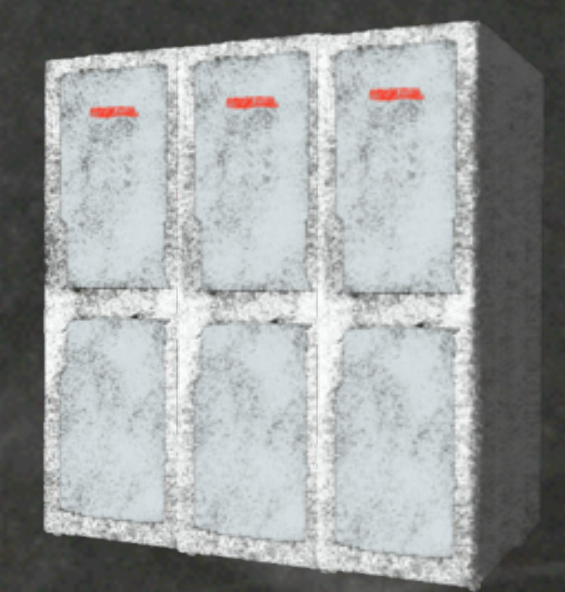


Specialized
supercomputing
resources

On-demand access to effectively
limitless resources

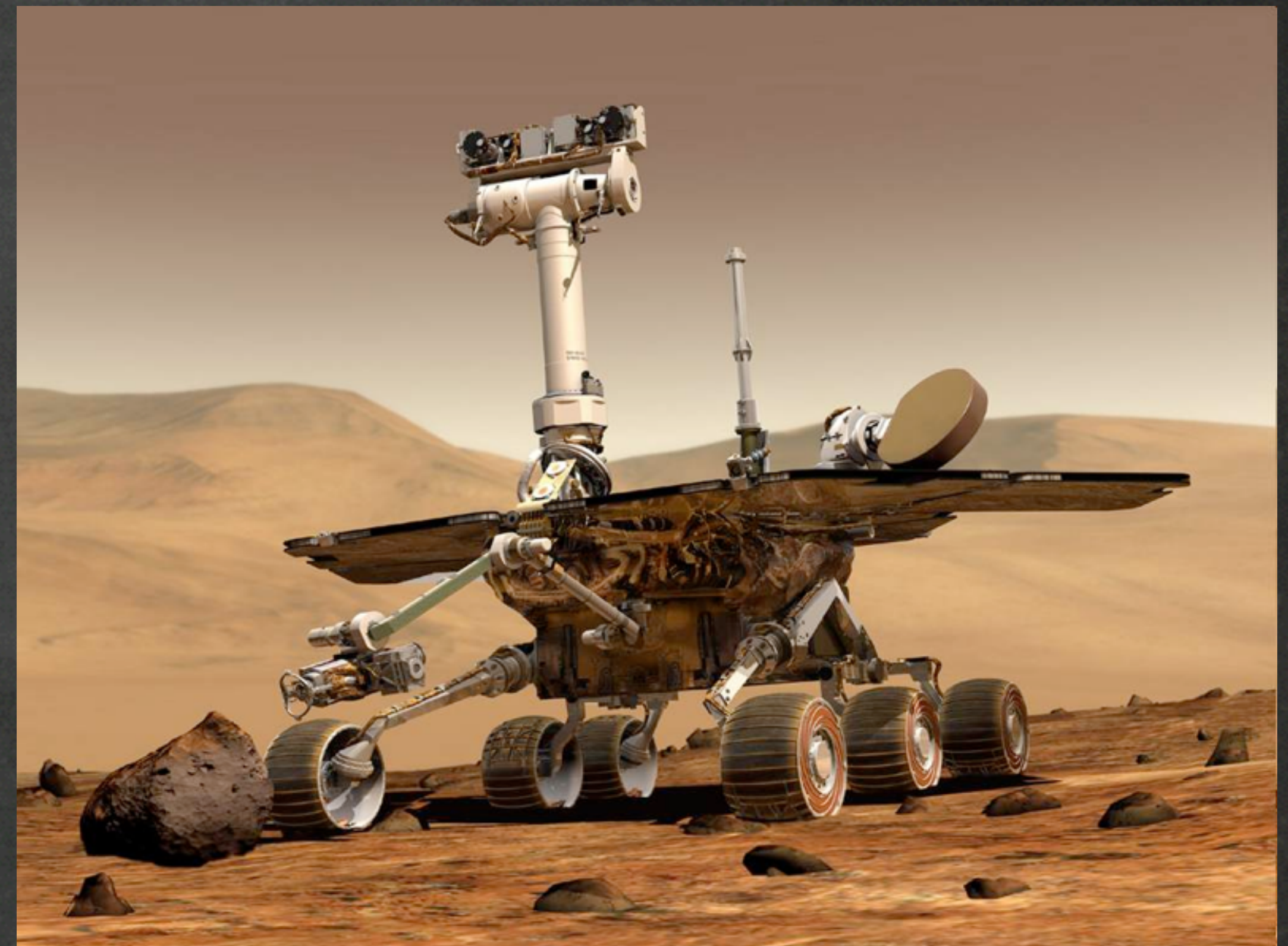
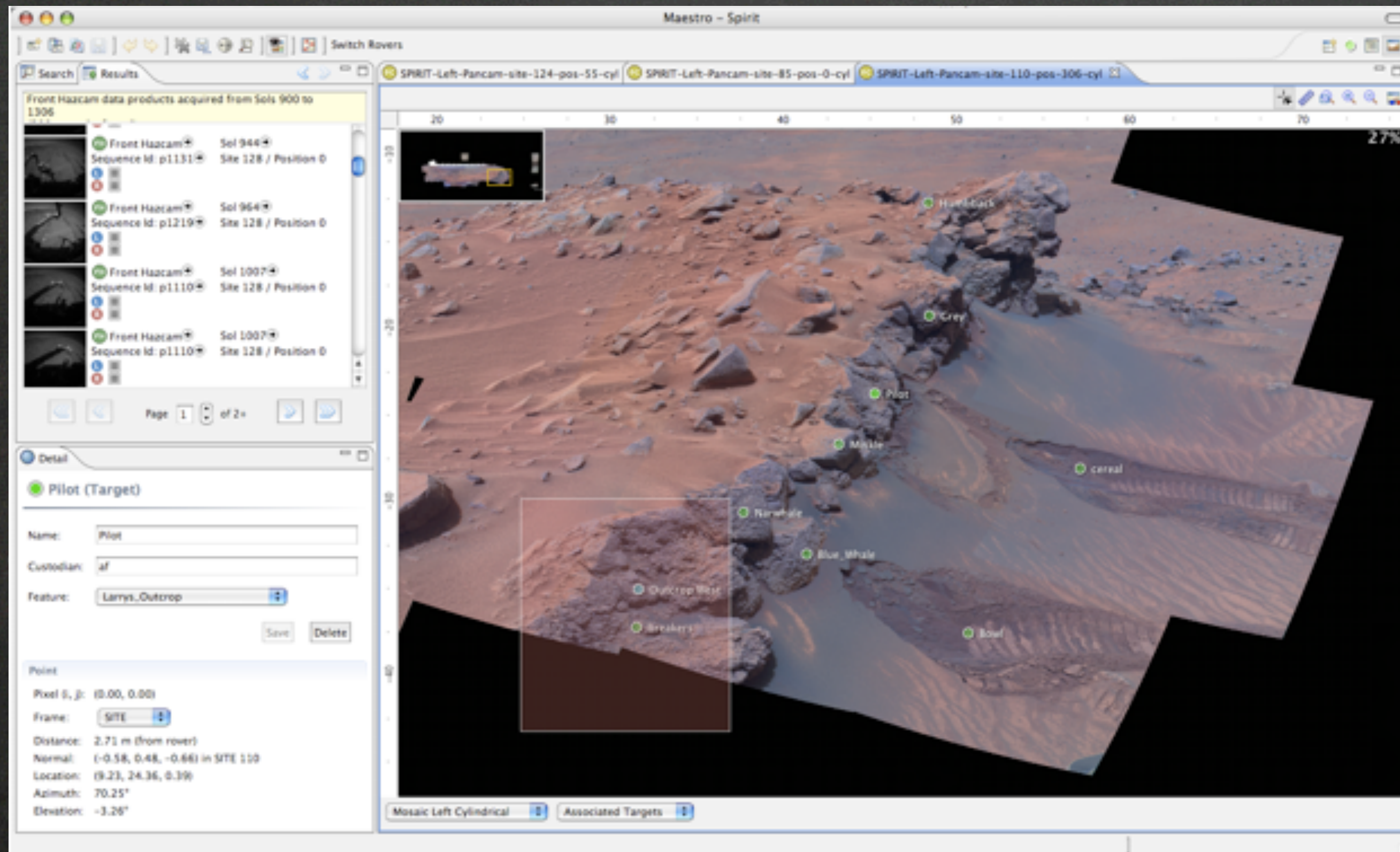


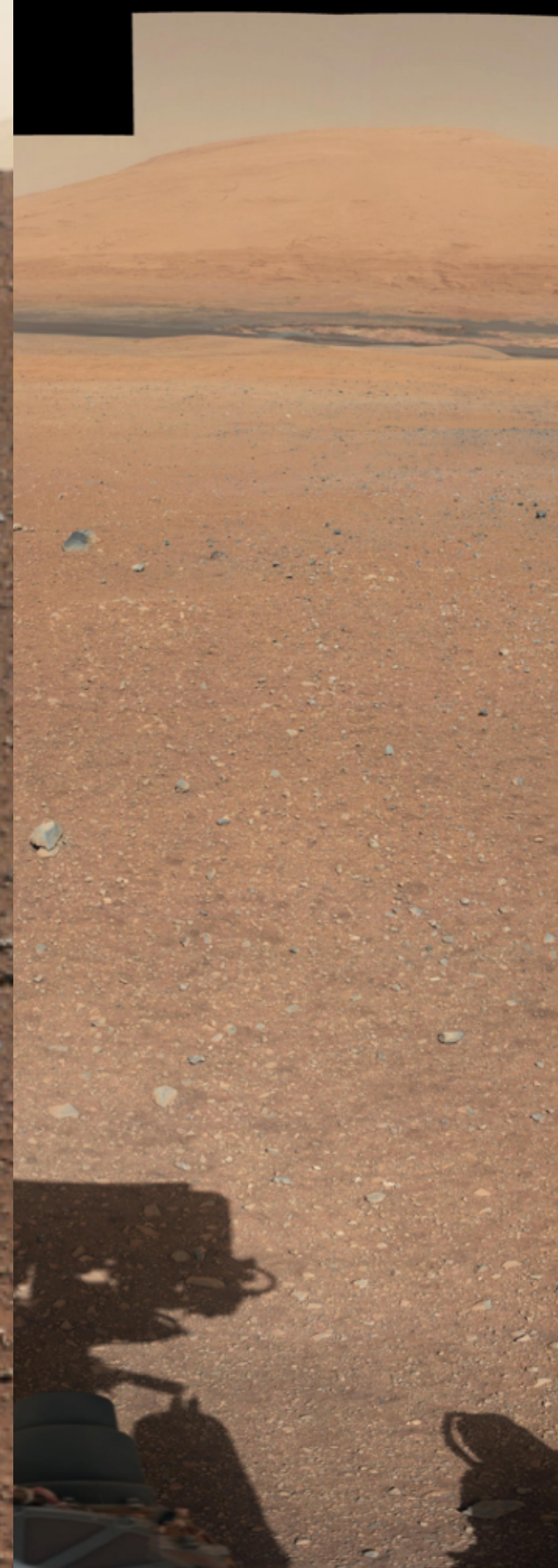
NASA
Researcher



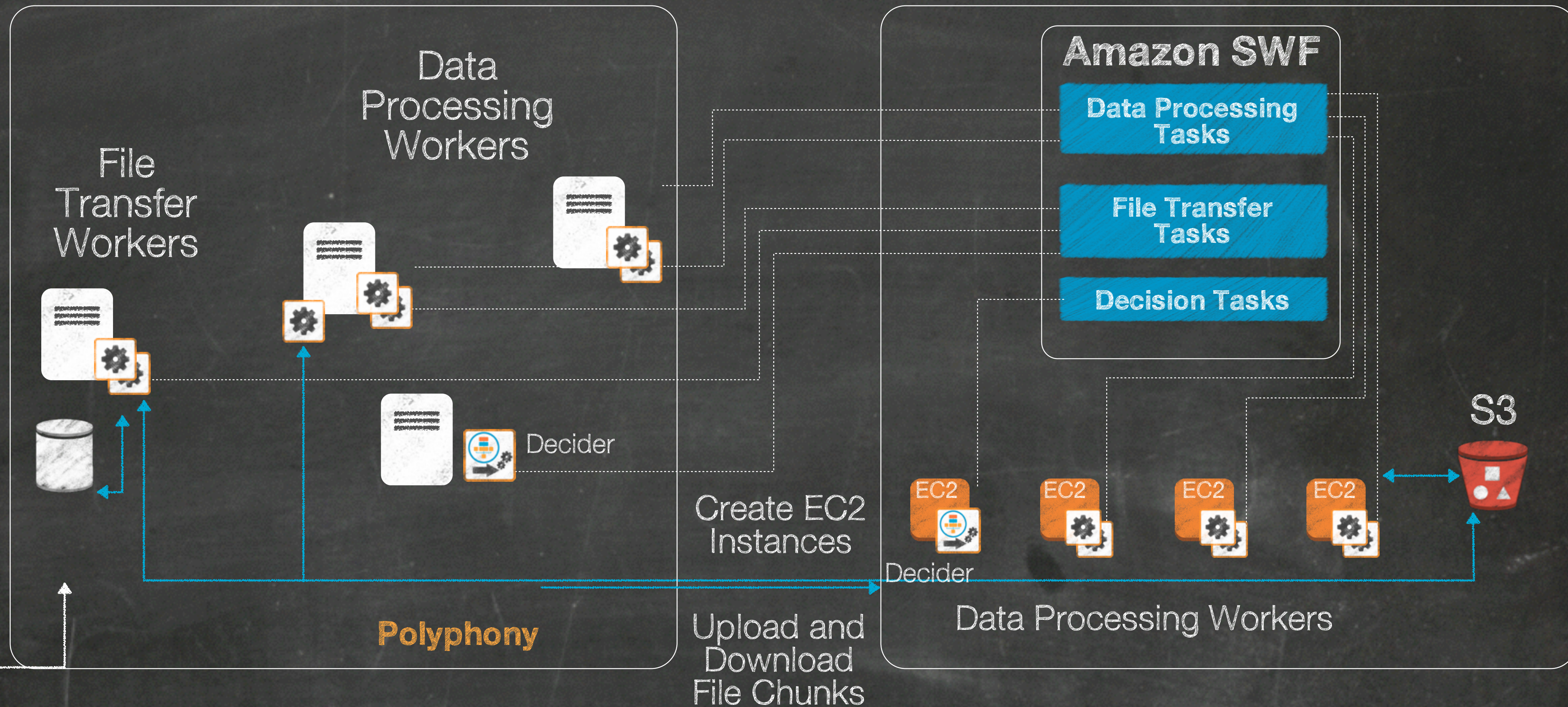
Small-scale
shared
compute

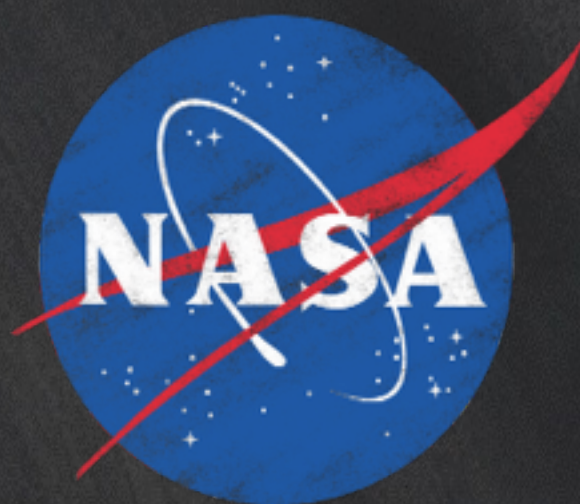
Mission-Critical Computing



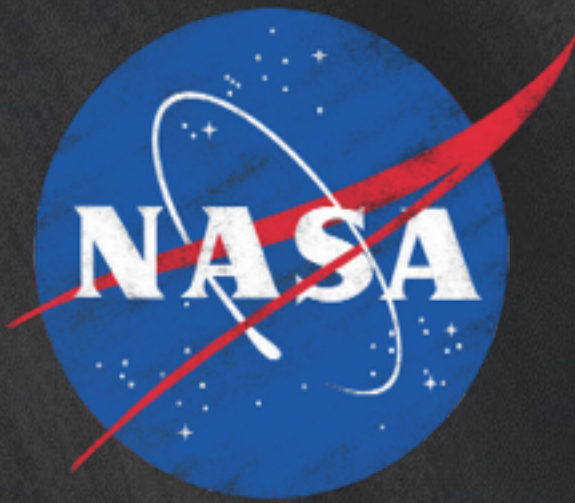


JPL Data Center





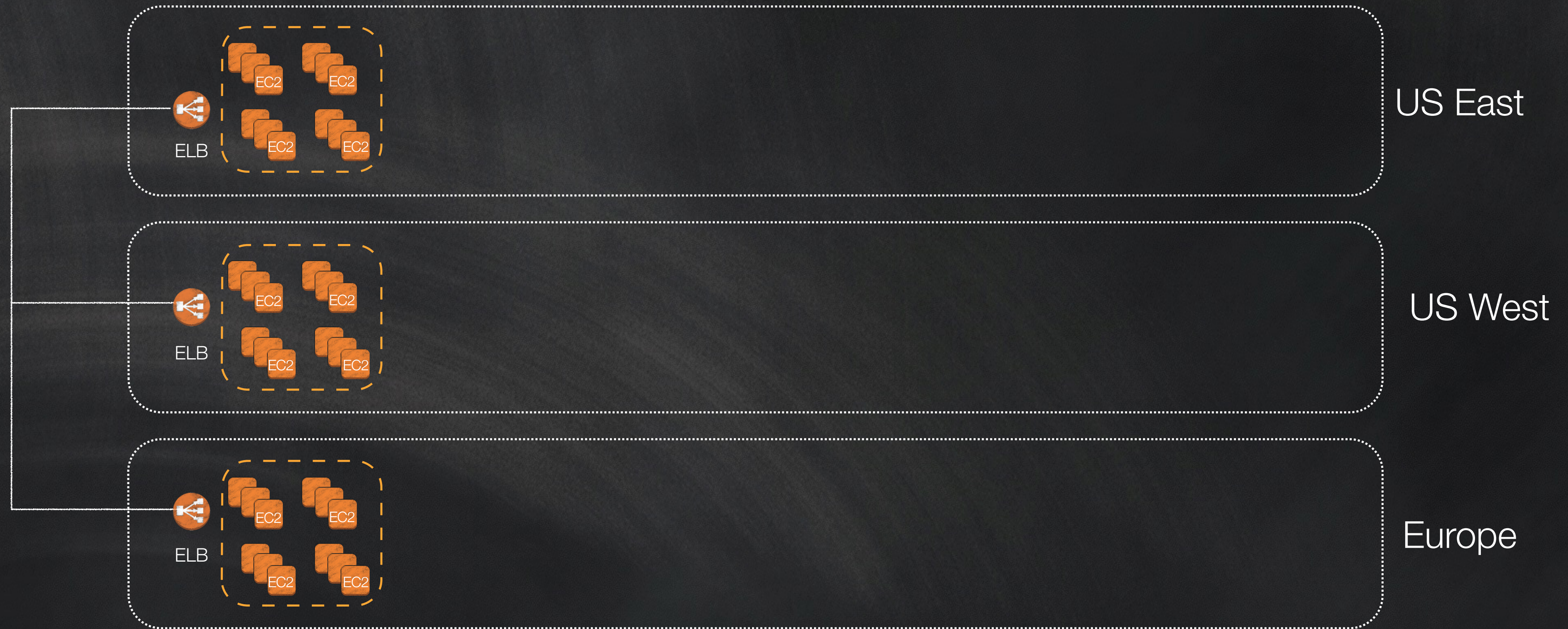
Curiosity live stream

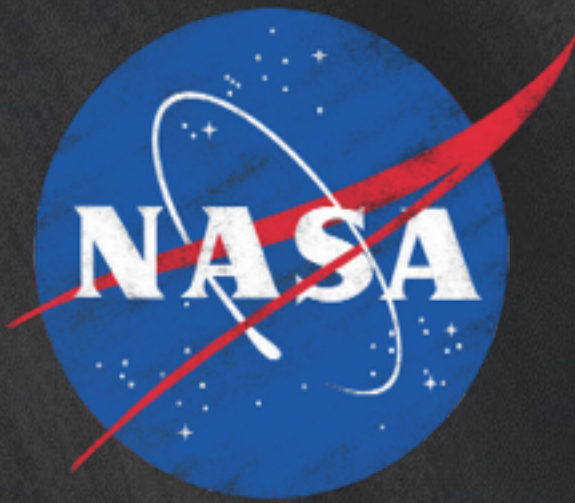


Curiosity live stream

Route 53: Multi-region weighted round-robin distribution

AWS Marketplace: Pre-configured Adobe Flash Media Server





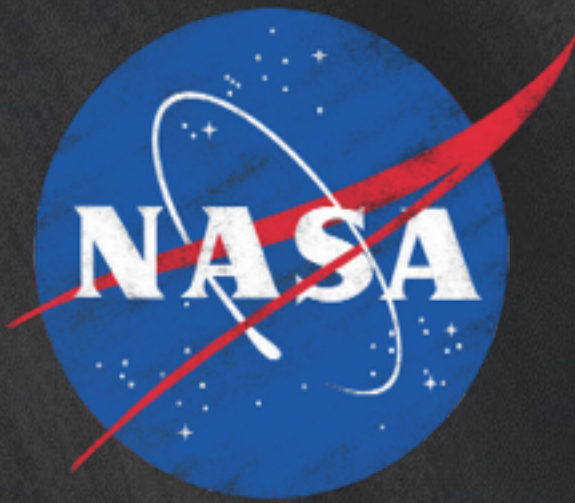
Curiosity live stream

Route 53: Multi-region weighted round-robin distribution

AWS Marketplace: Pre-configured Adobe Flash Media Server

CloudFormation: Quickly deploy repeatable units of streaming capacity





Curiosity live stream

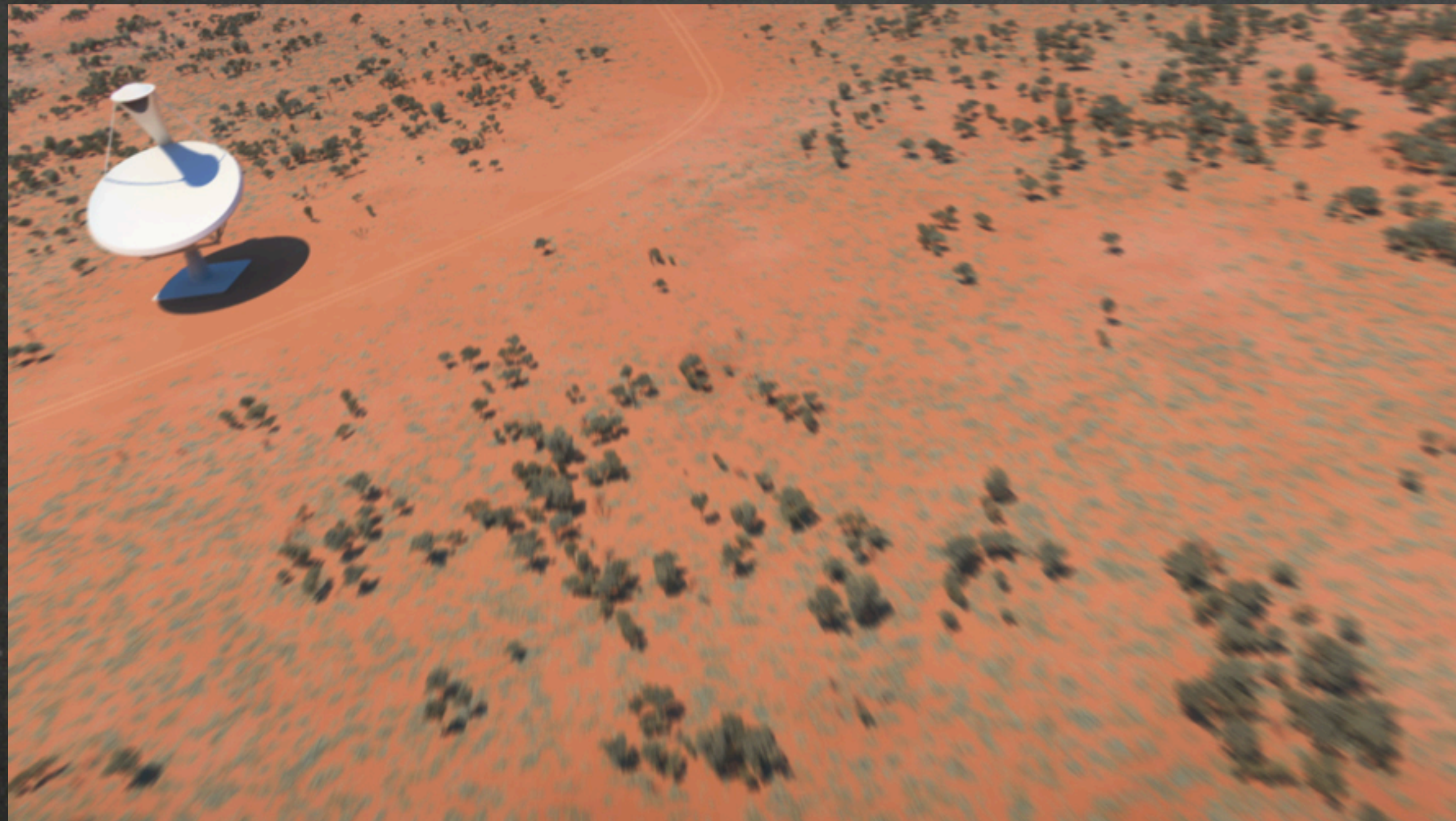
Route 53: Multi-region weighted round-robin distribution

AWS Marketplace: Pre-configured Adobe Flash Media Server

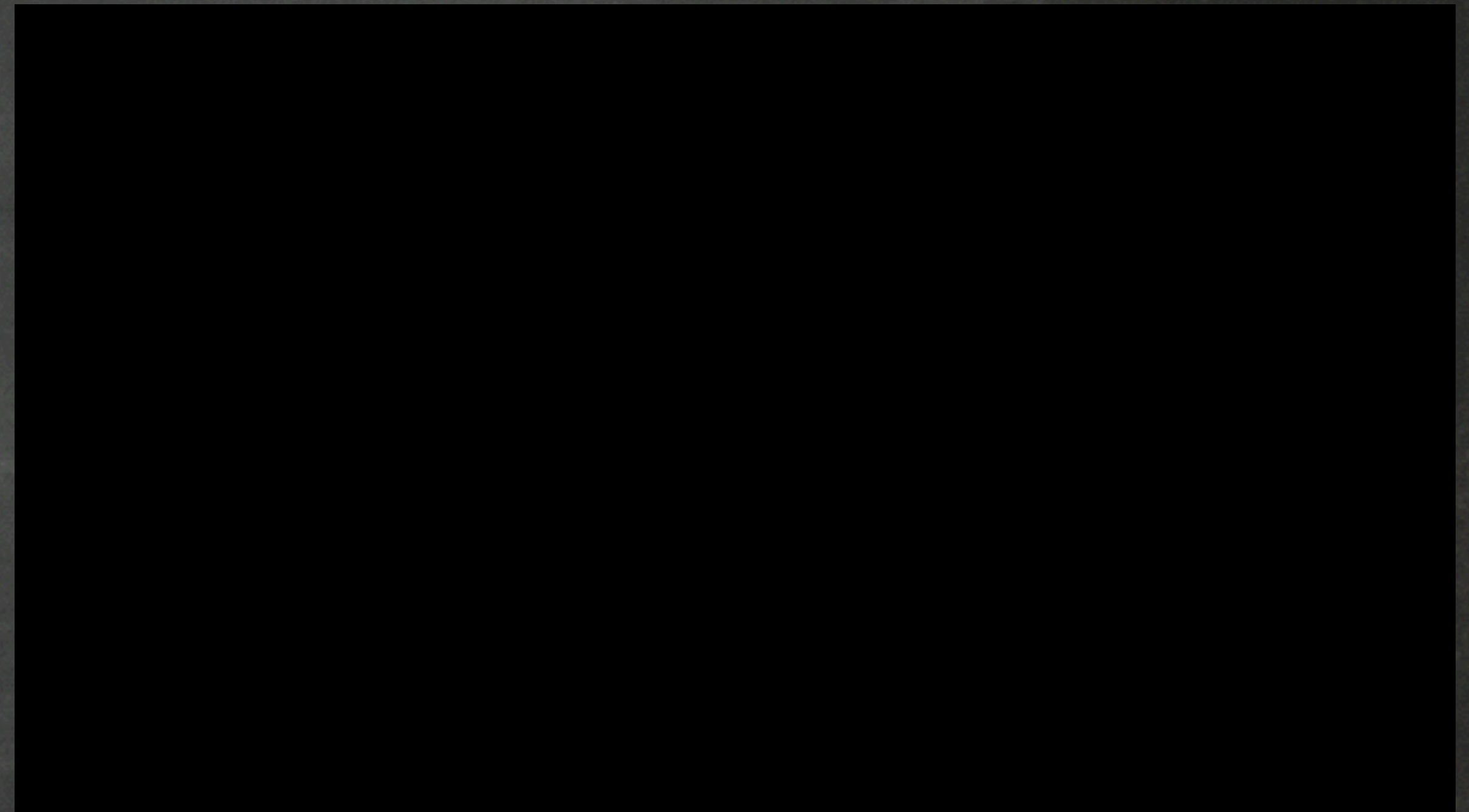
CloudFormation: Quickly deploy repeatable units of streaming capacity



The Square Kilometer Array



SKA Mid and Survey Dishes



SKA-low Dipoles

SCALING MWA / ASKAP TO SKA1

	MWA	ASKAP	SKA1-SURVEY	SKA1-LOW	SKA1-MID
Digitiser Output	0.12 Tbit/s 1x	69 Tbit/s 575x	184 Tbit/s 1533x	10 Tbit/s [3] 83x	17.1 Tbit/s [1] 143x
Input to Science Data Processor	3.2 Gbit/s 1x	20 Gbit/s 6.3x	37360 Gbit/s [2] 11675x	6736 Gbit/s [2] 2105x	26000 Gbit/s [2] 8125x
Archived Science Data Products	3 PB/year (25% duty cycle) [4]	5 PB/year	Exabytes per year	Exabytes per year	Exabytes per year

[1] From SADT Consortium Technical Development Plan (SKA-TEL.SADT-PROP_TECH-RED-001)

[2] From SKA1 System Baseline Design (SKA-TEL-SKO-DD-001)

[3] From LFAA Technical Description (AADC-TEL.LFAA.SE.MGT-AADC-PL-002)

[4] MWA is archiving correlator output data, not science data processor output data



www.theSkyNet.org

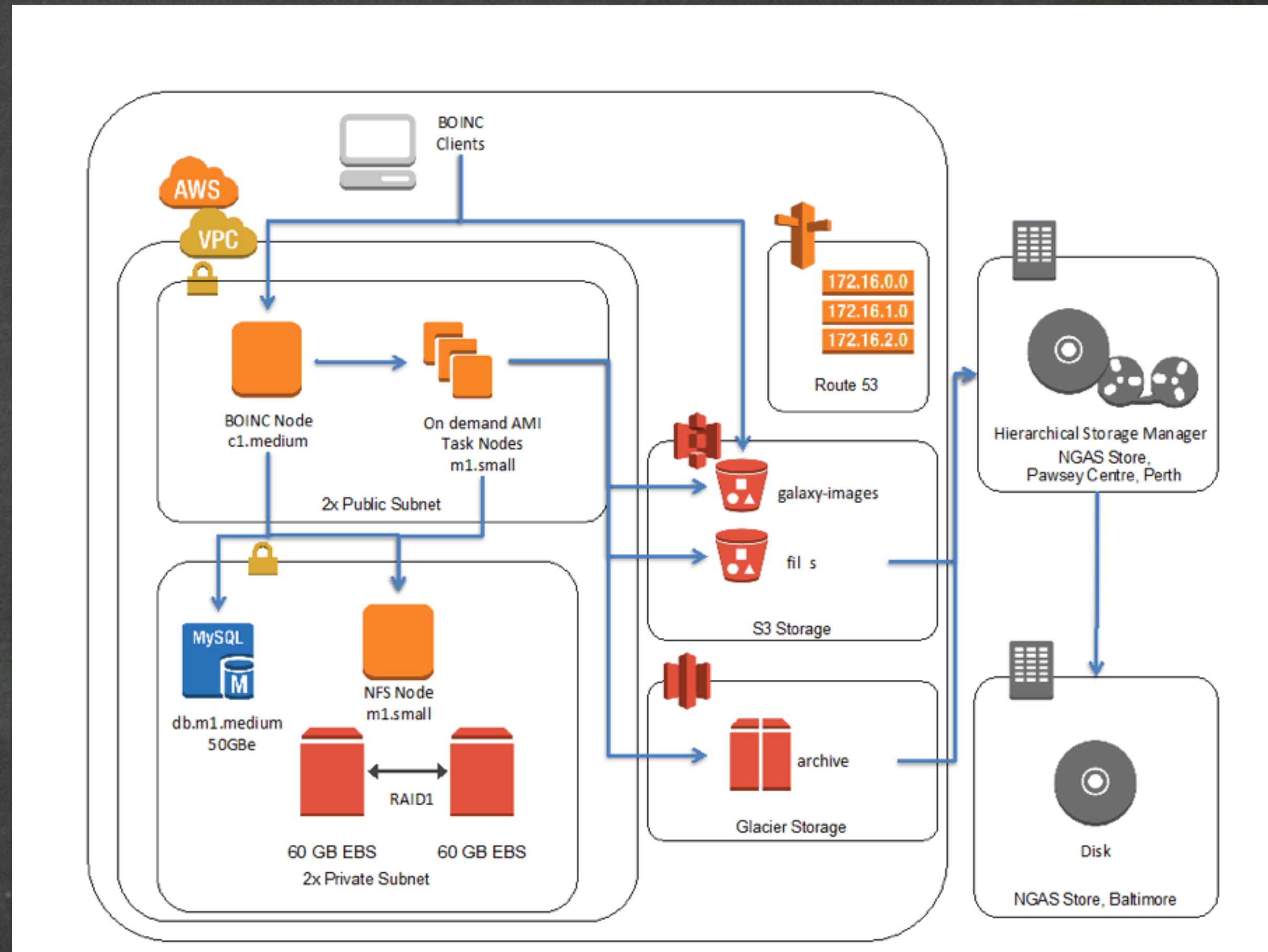
The Australian Square Kilometre Array Pathfinder (ASKAP) in the
Murchison Radio-astronomy Observatory (MRO), Western Australia.
Image courtesy of the Western Australian Department of Commerce



SkyNet - ICRAR EPO System for the SKA



International Centre for Radio Astronomy Research

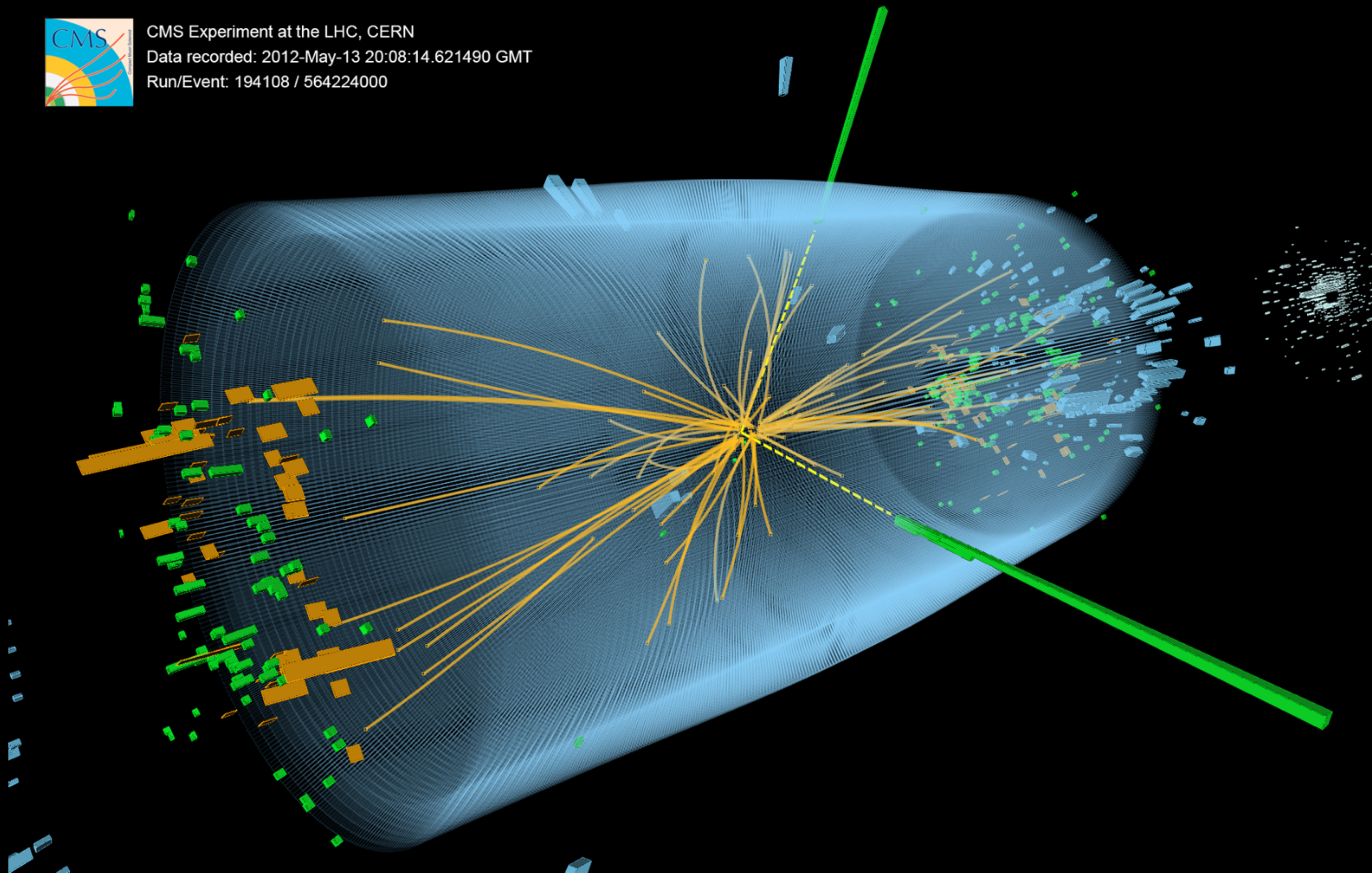


ICRAR SkyNet Architecture

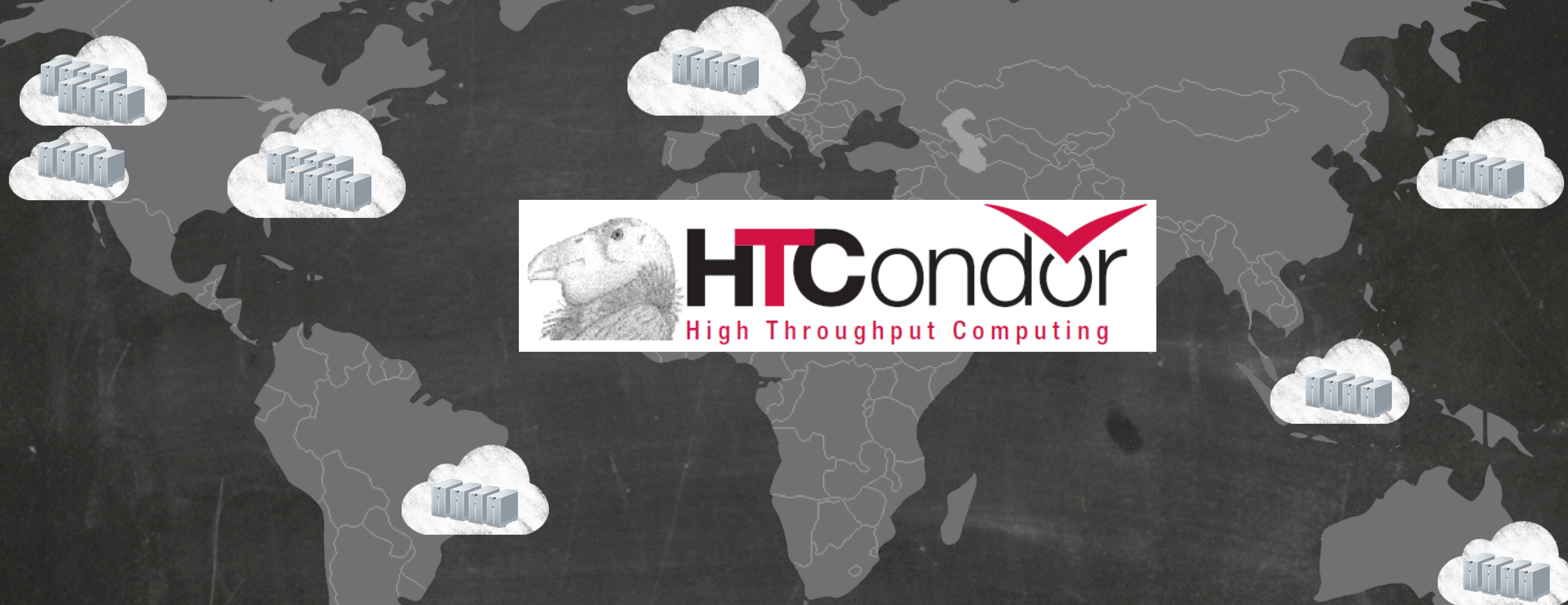
<http://aws.amazon.com/solutions/case-studies/icrar/>



CMS Experiment at the LHC, CERN
Data recorded: 2012-May-13 20:08:14.621490 GMT
Run/Event: 194108 / 564224000



Globally Distributed Compute for LHC on Amazon EC2 Spot



<http://www.hep.wisc.edu/~dan/talks/EC2SpotForCMS.pdf>

DIRAC @ Belle II

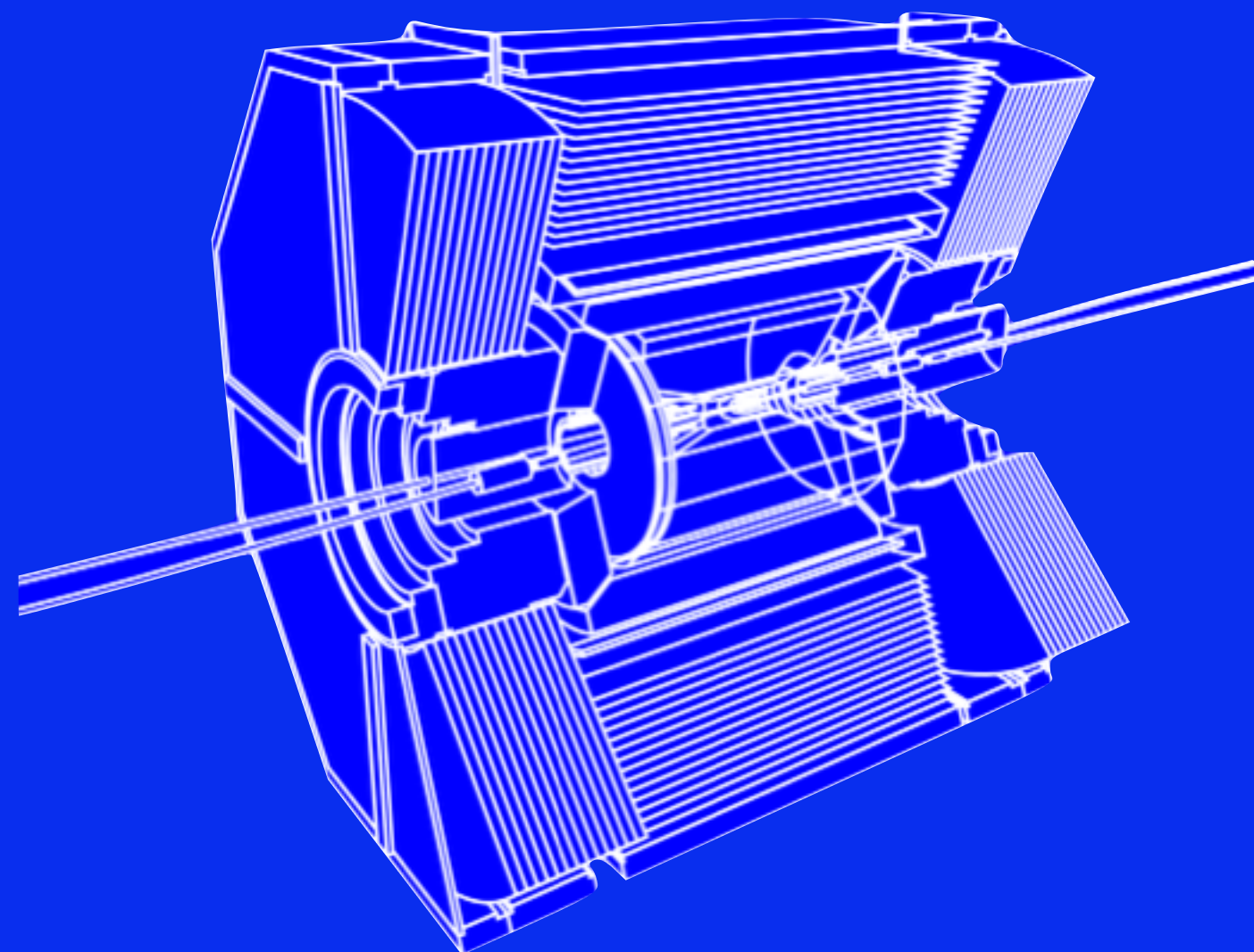
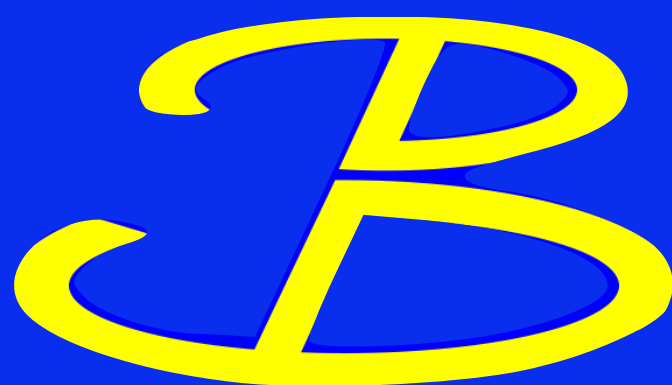


KEKB: 1 ab^{-1}

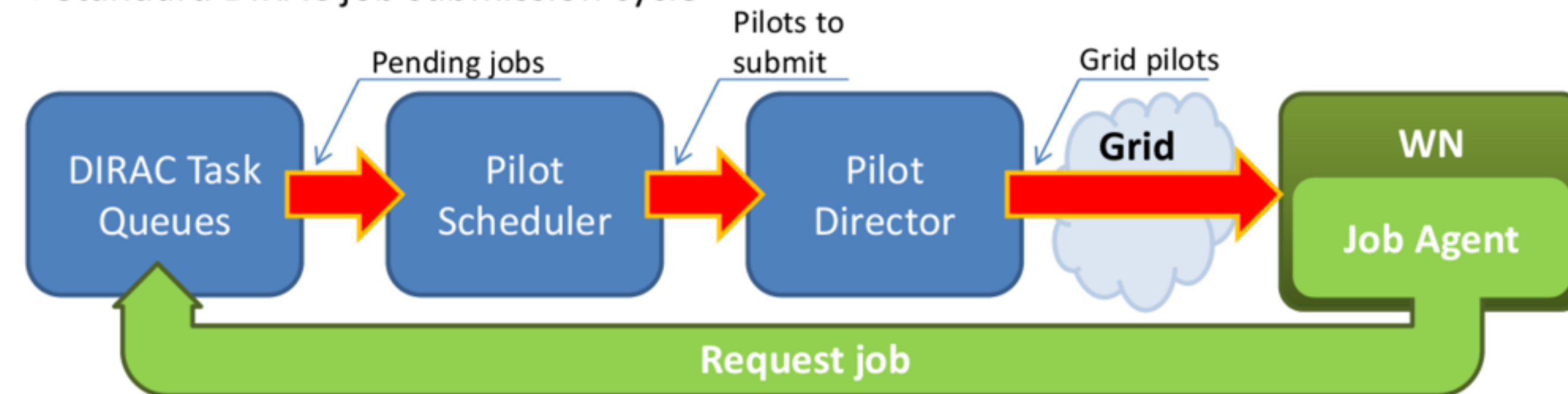


SuperKEKB: 50 ab^{-1}

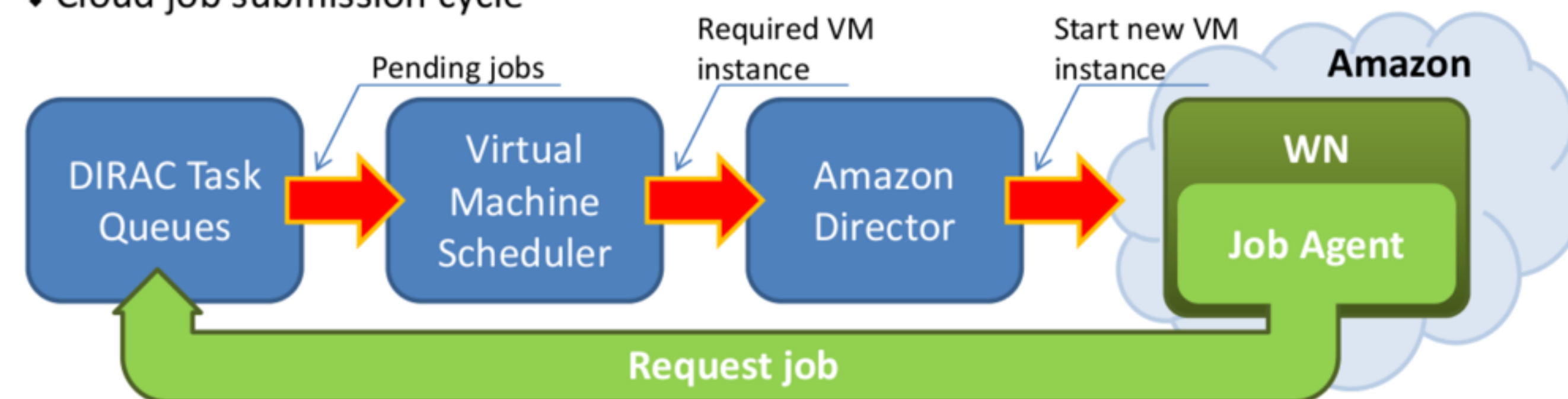




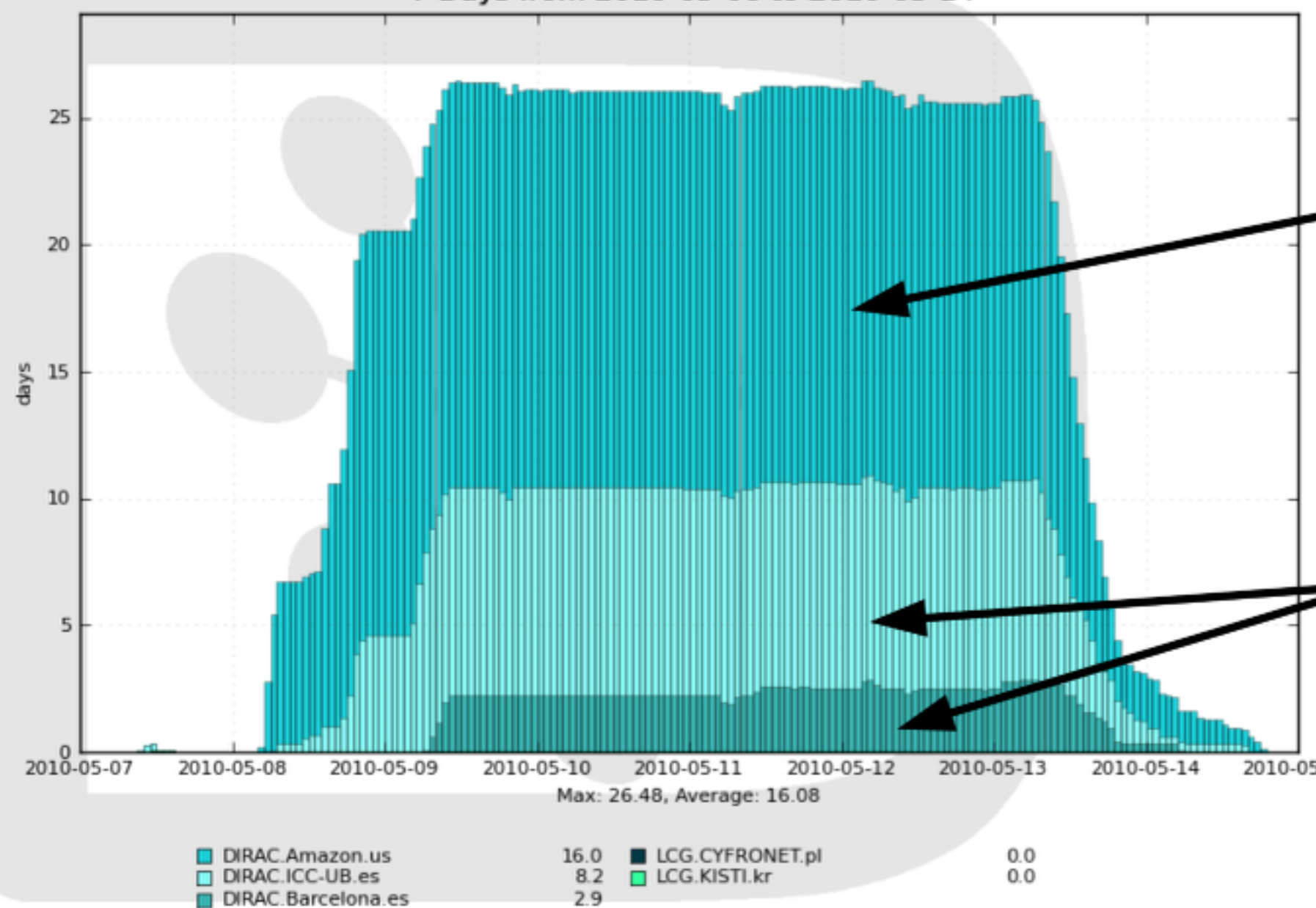
❖ Standard DIRAC job submission cycle



❖ Cloud job submission cycle



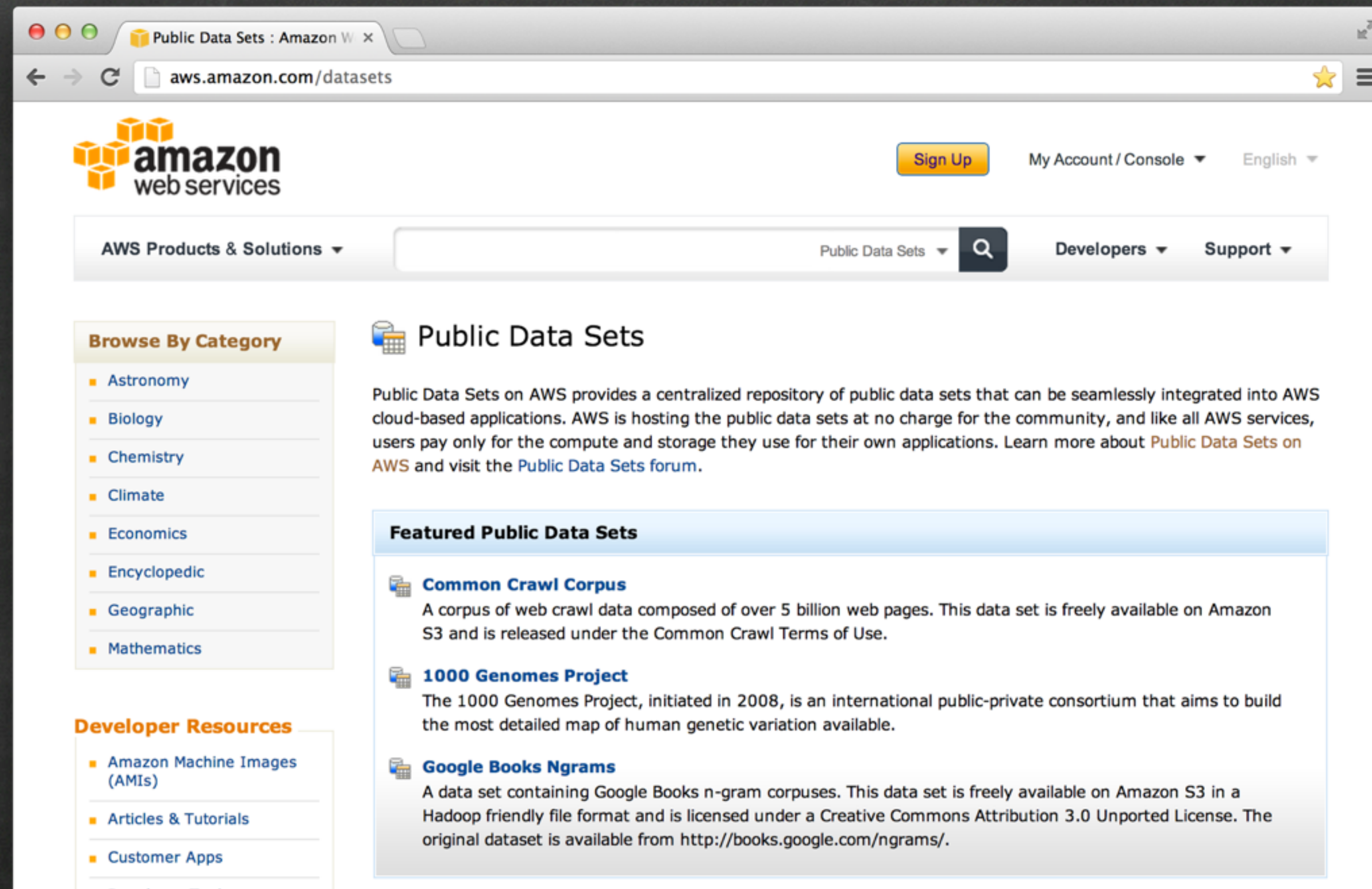
CPU days consumed by Site / hour
7 Days from 2010-05-06 to 2010-05-14



Hybrid Computation Model

- <- Belle on AWS Price/Performance Benchmark
- 170M events (3.6 TB) produced in 6 days
- Amazon Spot Instances -> \$0.20 / 10k events (May, 2010 pricing)

AWS Public Data Sets



The screenshot shows the AWS Public Data Sets page in a web browser. The browser's address bar displays 'aws.amazon.com/datasets'. The page features the AWS logo and navigation links such as 'Sign Up', 'My Account / Console', and 'English'. A search bar is present with 'Public Data Sets' entered. On the left, a 'Browse By Category' sidebar lists various fields: Astronomy, Biology, Chemistry, Climate, Economics, Encyclopedic, Geographic, and Mathematics. Below this, 'Developer Resources' includes links for Amazon Machine Images (AMIs), Articles & Tutorials, Customer Apps, and Developer Tools. The main content area is titled 'Public Data Sets' and includes a descriptive paragraph about the service. It also features a 'Featured Public Data Sets' section with three entries: 'Common Crawl Corpus', '1000 Genomes Project', and 'Google Books Ngrams', each with a brief description of the data set.

Public Data Sets : Amazon W x

aws.amazon.com/datasets

amazon web services

Sign Up My Account / Console English

AWS Products & Solutions Public Data Sets Developers Support

Browse By Category

- Astronomy
- Biology
- Chemistry
- Climate
- Economics
- Encyclopedic
- Geographic
- Mathematics

Developer Resources

- Amazon Machine Images (AMIs)
- Articles & Tutorials
- Customer Apps
- Developer Tools

Public Data Sets

Public Data Sets on AWS provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications. AWS is hosting the public data sets at no charge for the community, and like all AWS services, users pay only for the compute and storage they use for their own applications. Learn more about [Public Data Sets on AWS](#) and visit the [Public Data Sets forum](#).

Featured Public Data Sets

- Common Crawl Corpus**
A corpus of web crawl data composed of over 5 billion web pages. This data set is freely available on Amazon S3 and is released under the Common Crawl Terms of Use.
- 1000 Genomes Project**
The 1000 Genomes Project, initiated in 2008, is an international public-private consortium that aims to build the most detailed map of human genetic variation available.
- Google Books Ngrams**
A data set containing Google Books n-gram corpora. This data set is freely available on Amazon S3 in a Hadoop friendly file format and is licensed under a Creative Commons Attribution 3.0 Unported License. The original dataset is available from <http://books.google.com/ngrams/>.

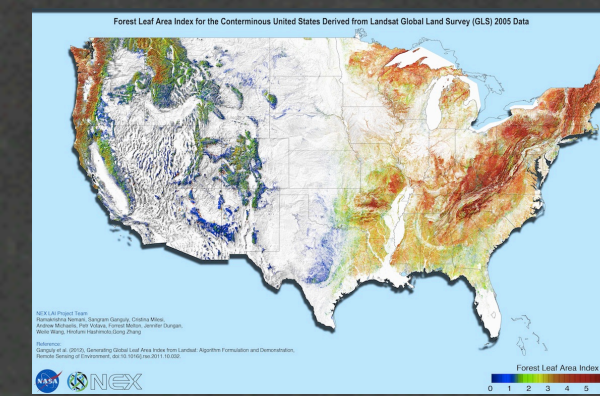
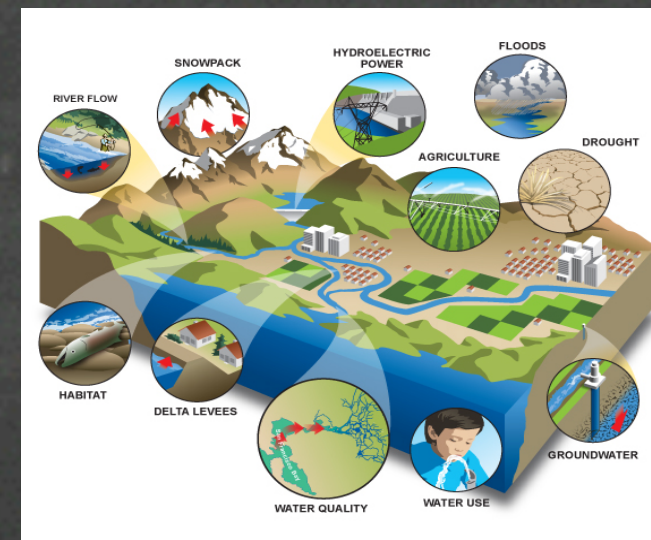
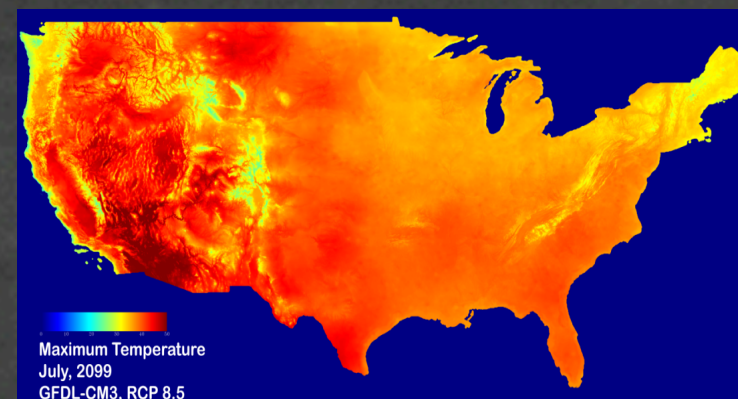
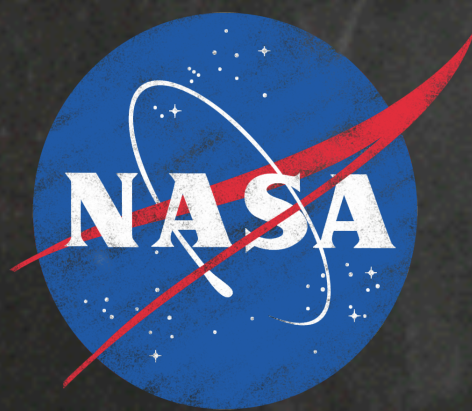
1000 Genomes
A Deep Catalog of Human Genetic Variation



[AWS.amazon.com/datasets](https://aws.amazon.com/datasets)

AWS and the NASA Earth eXchange (NEX)

- National Climate Assessment datasets hosted on AWS
- Machine images, tutorials and hosted workshops provided by NASA
- Data and Software now available to those without @nasa.gov email addresses
- Enables crowd-sourced citizen science applications like those found on the Zooniverse



Climate Researchers



Multi-Spectrum Atlas of the Galactic Plane

- Collaboration between AWS, Caltech/IPAC and USC/ISI
- All images are publicly accessible via direct download and VAO APIs
- 16 wavelength infrared atlas spanning $1\mu\text{m}$ to $70\mu\text{m}$
- Datasets from GLIMPSE and MIPS GAL, 2MASS, MSX, WISE
- Spatial sampling of 1 arcsec with $\pm 180^\circ$ longitude and $\pm 20^\circ$ latitude
- Mosaics generated by Caltech's Montage (<http://montage.ipac.caltech.edu>)
- Compute resources coordinated by USC's Pegasus (<http://pegasus.isi.edu/>)

Cancer Research with AWS

1000 Genomes Project and AWS

RELATED LINKS

Life Sciences on AWS

HPC on AWS

1000 Genomes Project and AWS

The 1000 Genomes Project is an international research effort coordinated by a consortium of 75 companies and organizations to establish the most detailed catalogue of human genetic variation. The project has grown to 200 terabytes of genomic data including DNA sequenced from more than 1,700 individuals that researchers can now access on AWS for use in disease research. The 1000 Genomes Project aims to include the genomes of more than 2,662 individuals from 26 populations around the world, and the NIH will continue to add the remaining genome samples to the data collection this year.

The dataset containing the full genomic sequence of 1,700 individuals is now available to all via Amazon S3. The data can be found at: s3.amazonaws.com/1000genomes

Accessing 1000 Genomes Data

AWS is making the 1000 Genomes Project data publicly available to the community free of charge. Public Data Sets on AWS provide a centralized repository of public data hosted on Amazon Simple Storage Service (Amazon S3). The data can be seamlessly accessed from AWS services such as Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Elastic MapReduce (Amazon EMR), which provide organizations with the highly scalable compute resources needed to take advantage of these large data collections. AWS is storing the public data sets at no charge to the community. Researchers pay only for the additional AWS resources they need for further processing or analysis of the data. Learn more about [Public Data Sets on AWS](#).

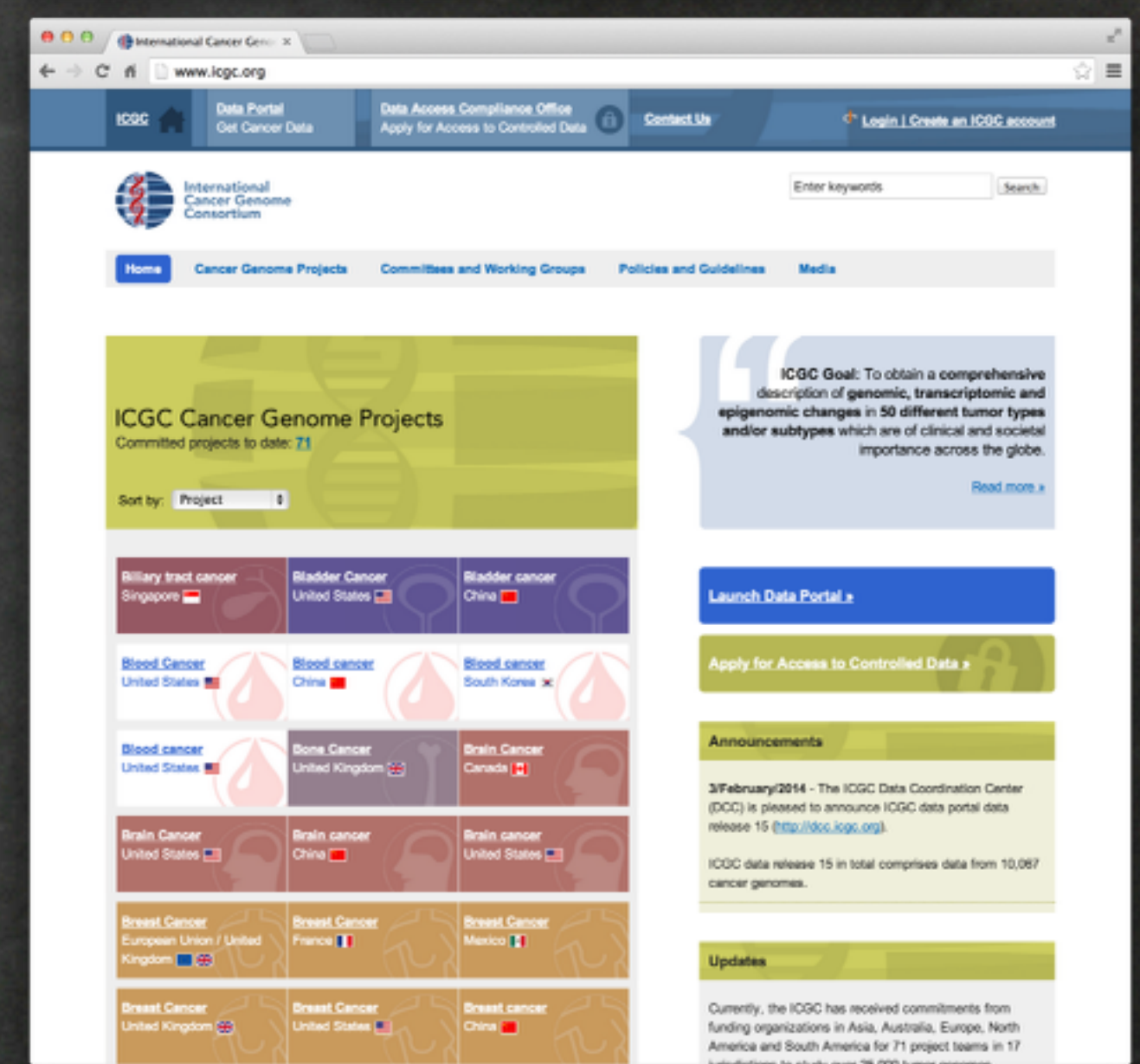
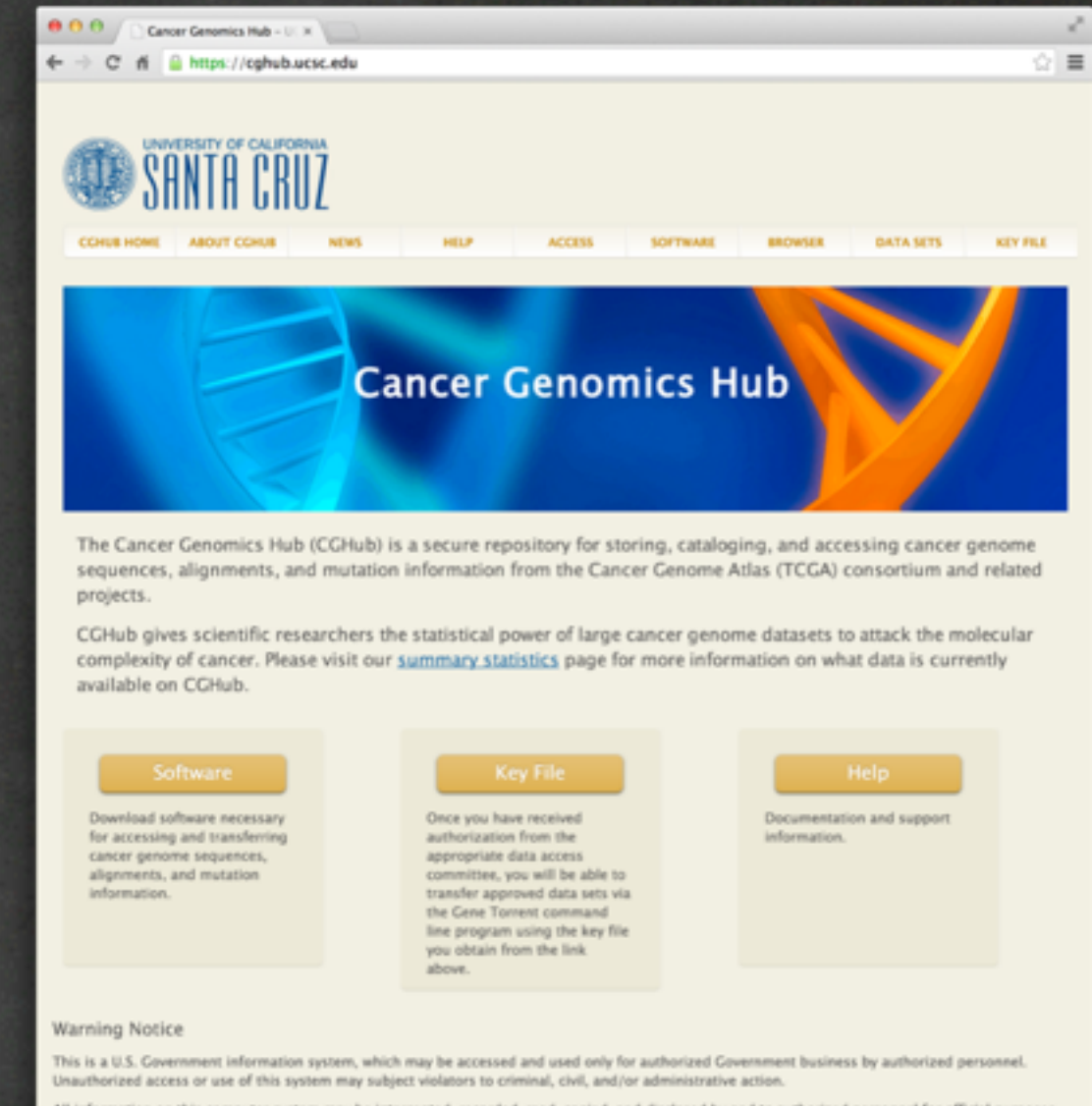
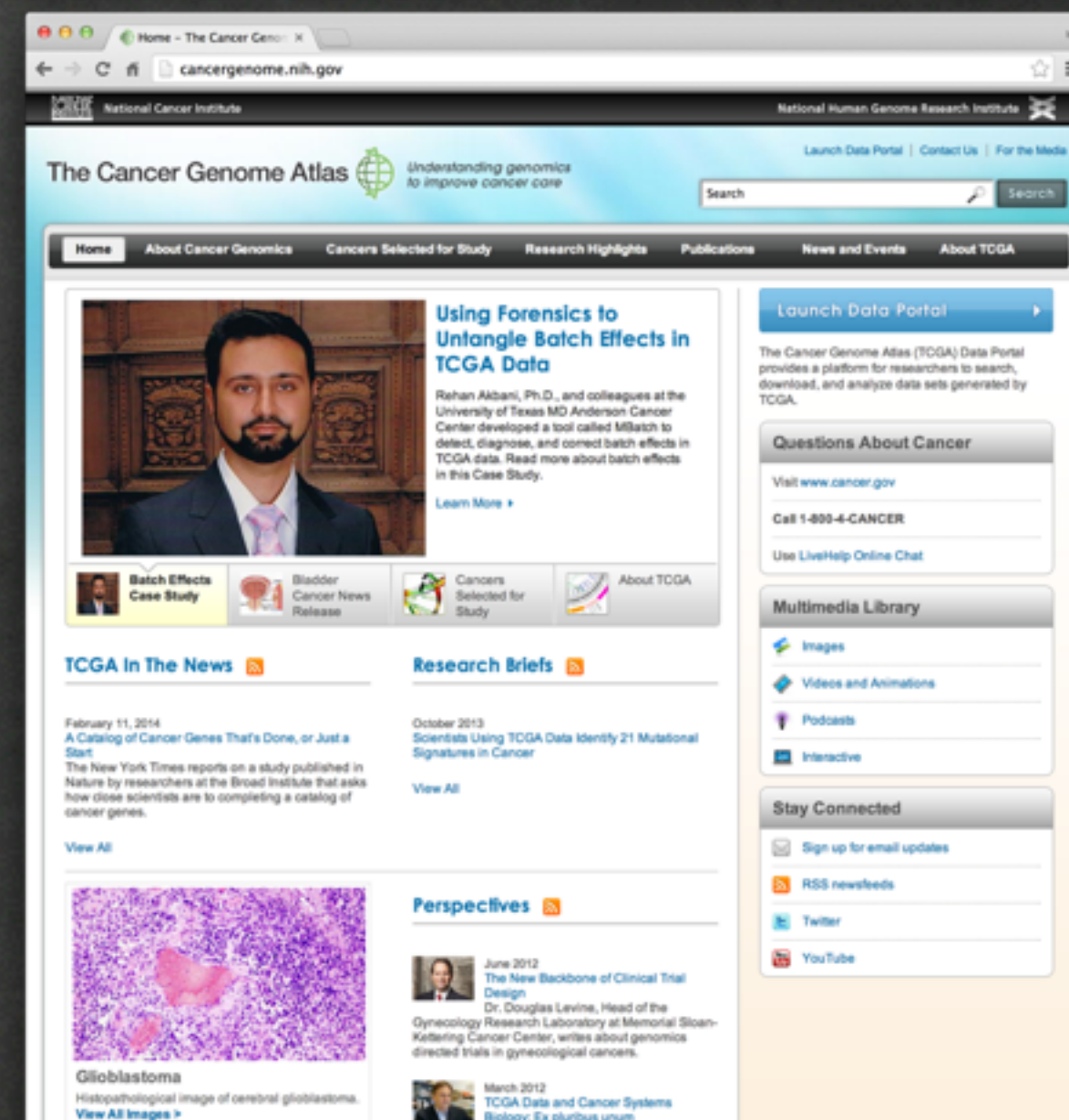
All 200 TB of the latest 1000 Genomes Project data is available in a [publicly available Amazon S3 bucket](#).

You can access the data via simple HTTP requests, or take advantage of the AWS SDKs in languages such as Ruby, Java, Python, .NET and PHP.

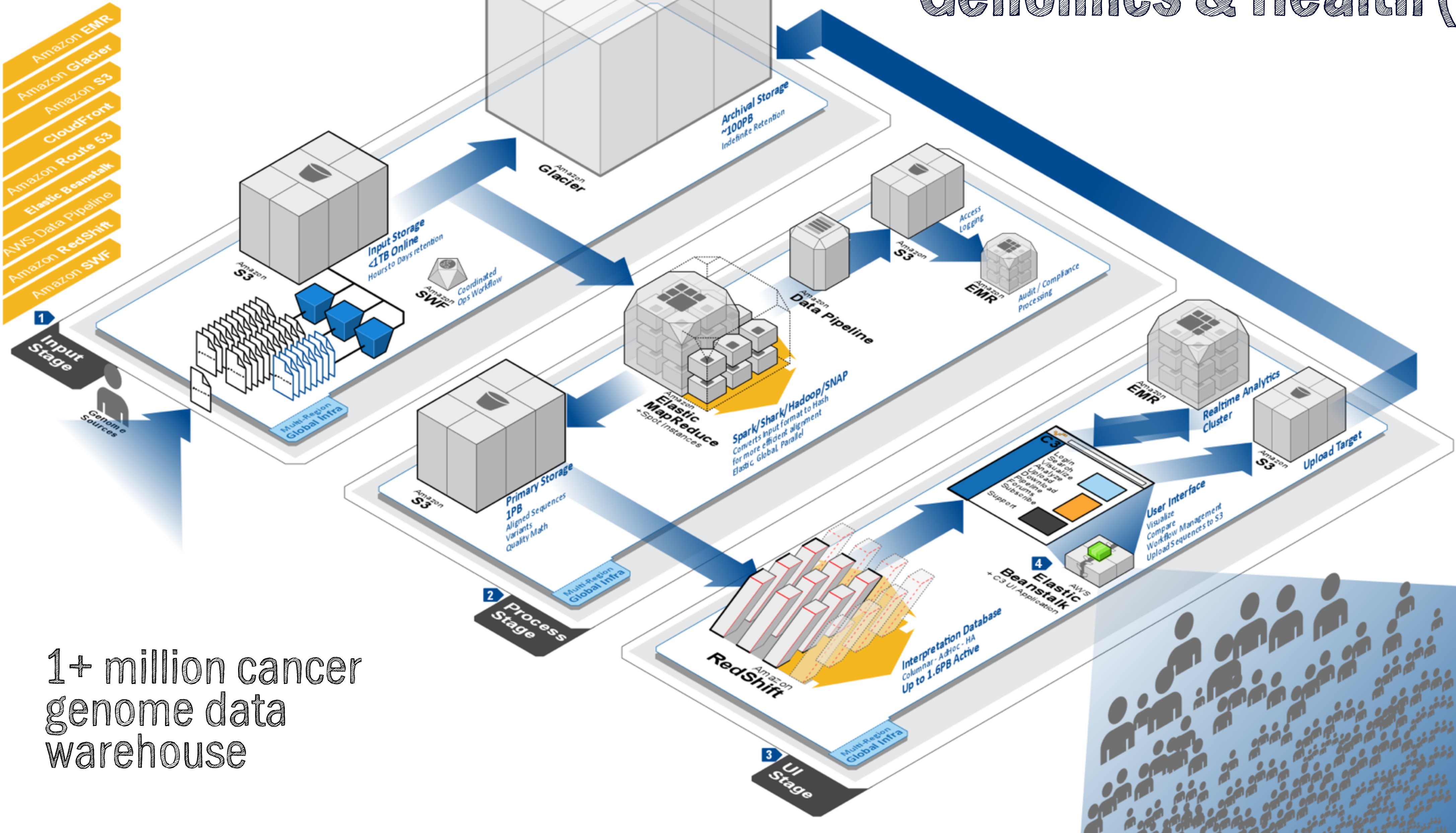
Analyzing 1000 Genomes Data

Researchers can use the Amazon EC2 utility computing service to dive into this data without the usual capital investment required to work with data at this scale. AWS also provides a number of [orchestration](#) and [automation](#) services to help teams make their research available to others to remix and reuse.

Making the data available via a bucket in Amazon S3 also means that customers can crunch the information using Hadoop via [Amazon Elastic MapReduce](#), and take advantage of the growing collection of tools for running bioinformatics job flows, such as [CloudBurst](#) and [Crossbow](#).



The Global Alliance for Genomics & Health (GA4GH)



1+ million cancer genome data warehouse

National Database for Autism Research

National Database for Autism Research (NDAR) cloud overview page. The page includes a navigation bar with links: Home, Query, Harmonization Tools, Cloud, Contribute, Request Access, Policy, Tutorials, About, FAQ, and a login button. The main content area is titled "Overview" and includes a "Get Started" button. The text describes the mission of NDAR: to make available all research data related to autism for reuse. It mentions that raw data is expected as are the results of each experiment, and that data collected across projects are aggregated and made available through the NDAR GUID. It also states that results from each experiment - often on the same subjects - can now be made available. In this way, separate experiments on genotypes and brain volumes can inform the research community on the tens of thousands of subjects now contained in NDAR. NDAR's cloud computation capability provides a framework in support of this infrastructure.

From

NDAR Data Packaging

NIH

Oracle Database

Accounts
Phenotype
Mappings to file location
Clinical/Imaging/Genomics data

Amazon Web Services

S3 Files

Imaging and -OMIC files in S3/
Glacier

Download/Copy

NDAR

To

Cloud Computational Model

NIH

Oracle Database

Accounts
Phenotype
Mappings to file location
Clinical/Imaging/Genomics data

Amazon Web Services

S3 Files

Imaging and -OMIC files in S3/
Glacier

RDS Database

Computation in Cloud

NITRC

Galaxy

Cloud BioLinux

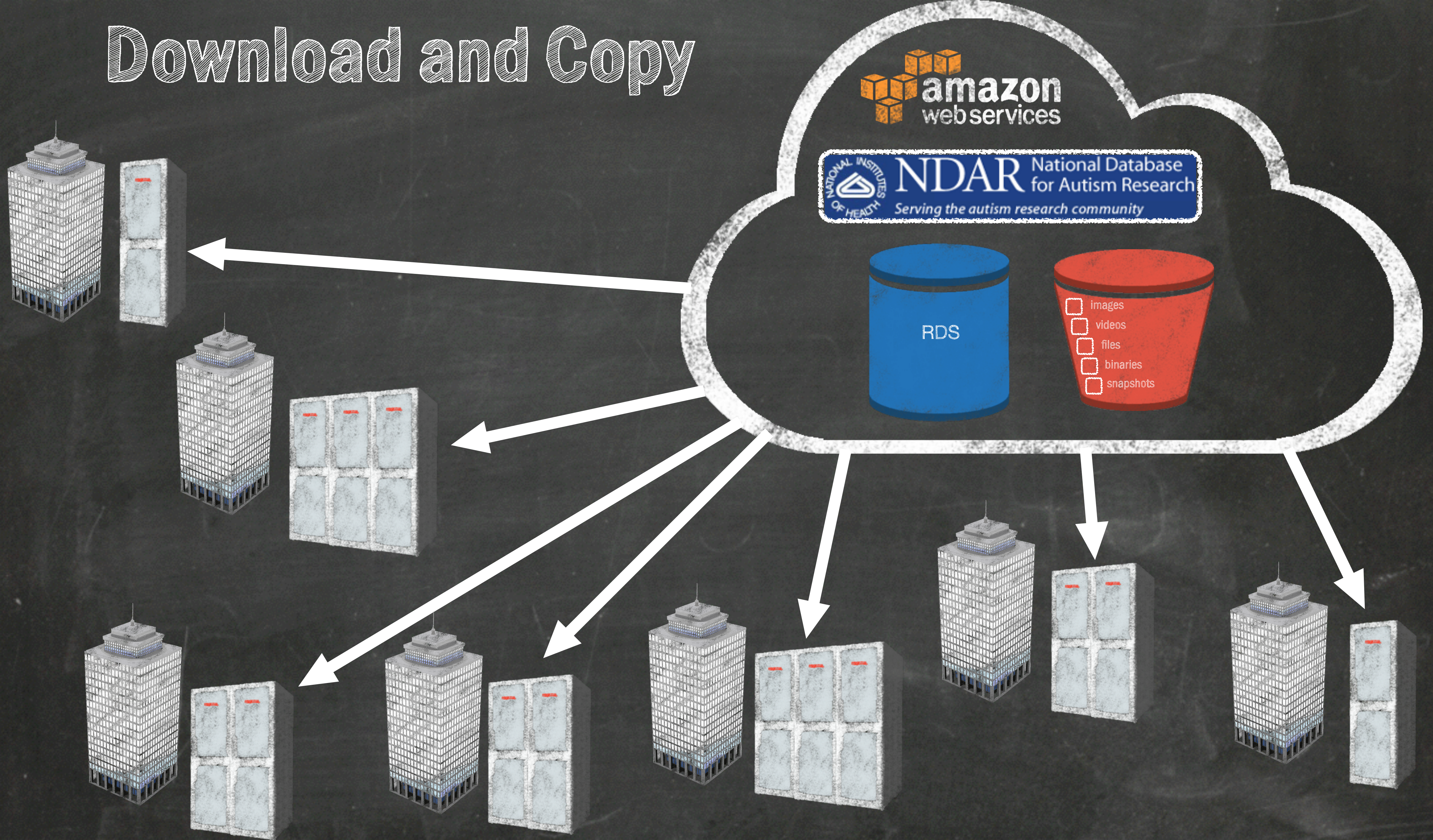
Contact Us | Privacy | Disclaimer | Accessibility | Site Map | FOIA | OIG | Government Warning Notice

http://ndar.nih.gov/cloud_overview.html

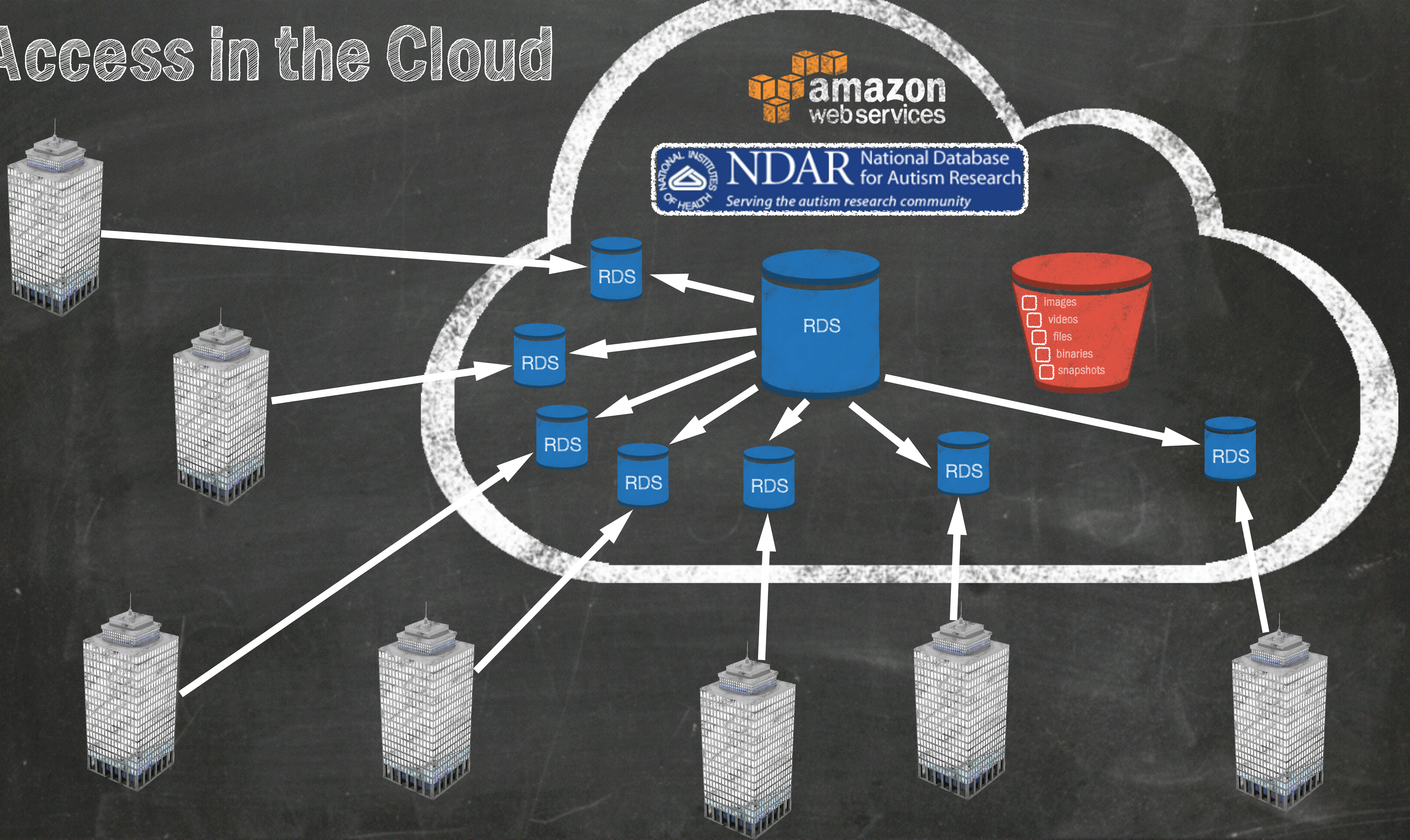
All autism research funded by NIMH must be publicly accessible

NDAR provides a web interface to query the aggregate data set

Download and Copy



Access in the Cloud



Computation in the Cloud



Where is this Going?

1. Researcher conducts experiment



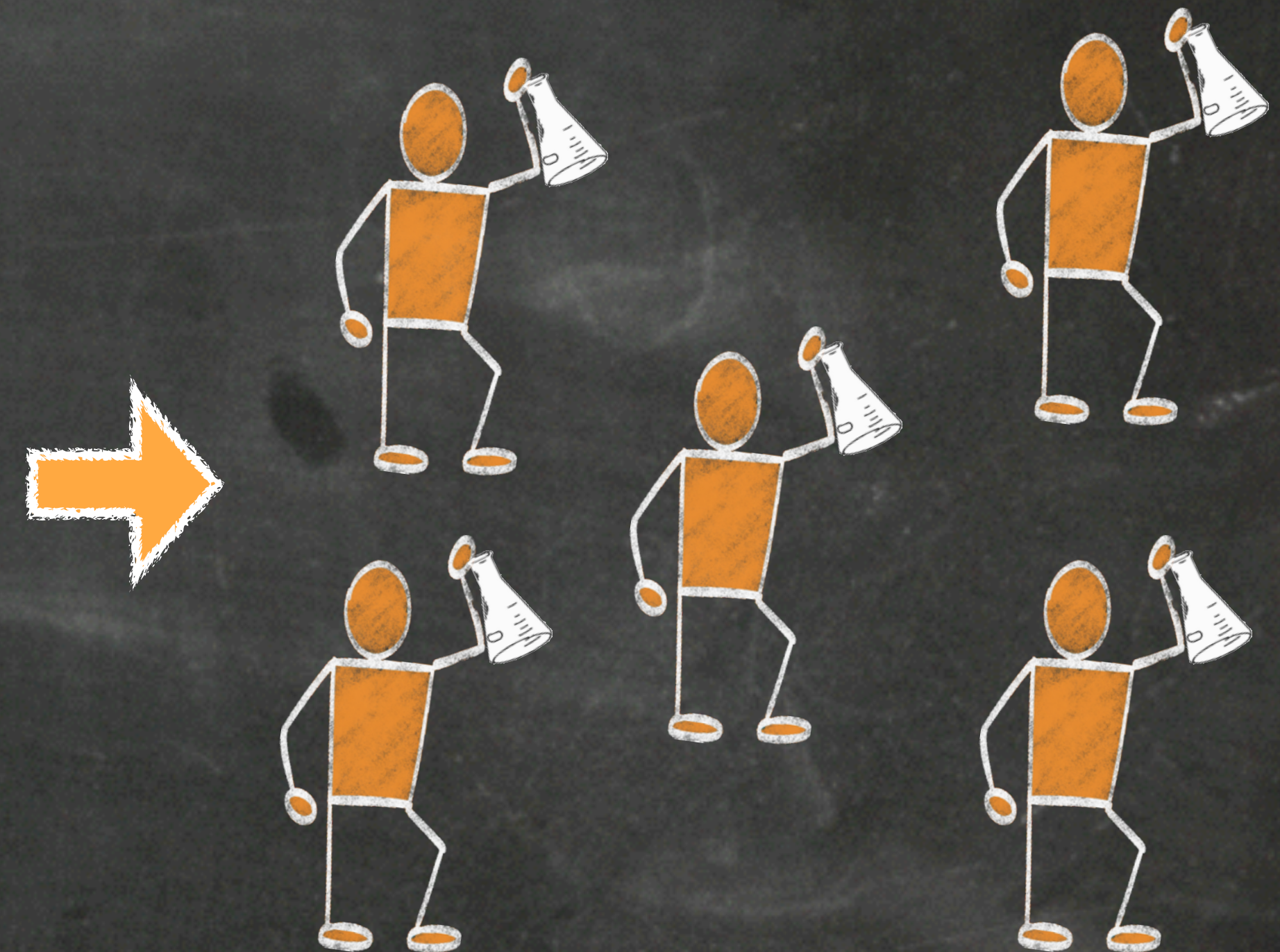
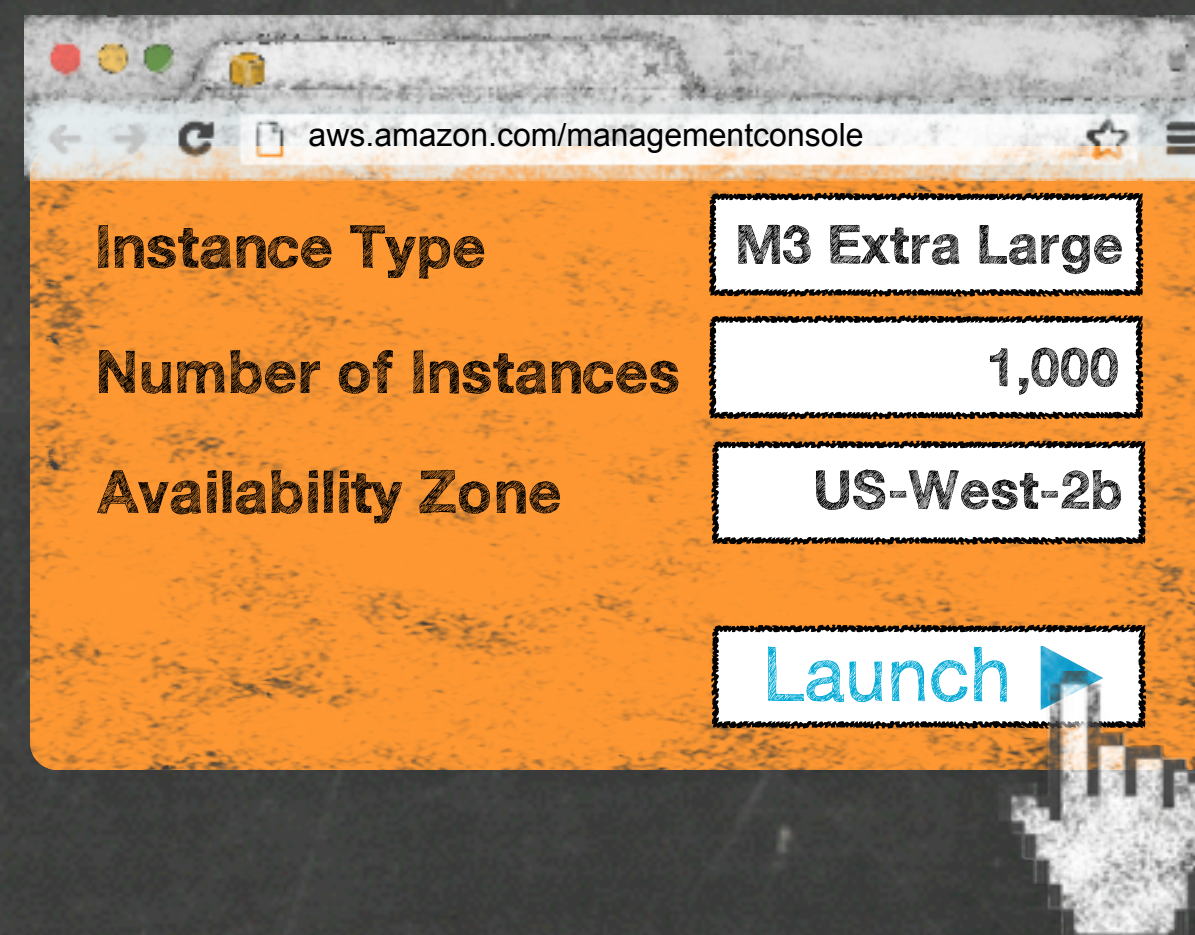
Where is this Going?

1. Researcher conducts experiment
2. Experimental data and results uploaded to the cloud along with reproducible machine images



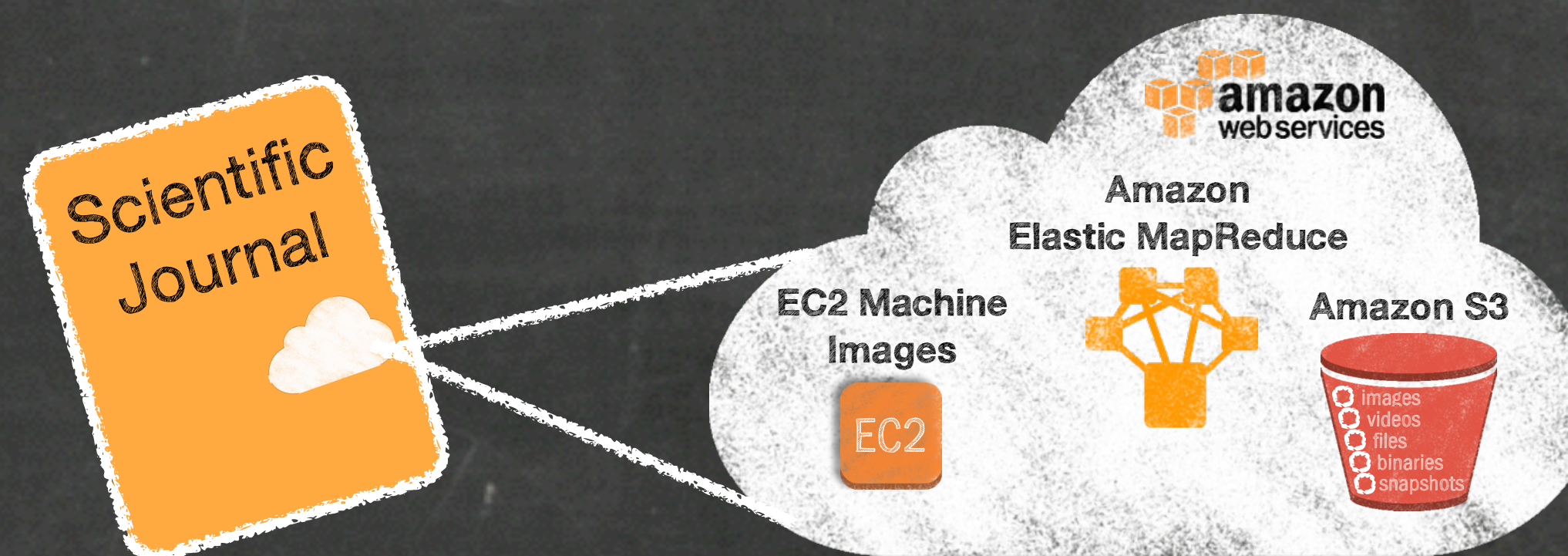
Where is this Going?

1. Researcher conducts experiment
2. Experimental data and results uploaded to the cloud along with reproducible machine images
3. Reviewers leverage cloud resources to reproduce and validate results.



Where is this Going?

1. Researcher conducts experiment
2. Experimental data and results uploaded to the cloud along with reproducible machine images
3. Reviewers leverage cloud resources to reproduce and validate results.
4. Results published in a peer-reviewed journal, including references (e.g. DOIs) to cloud data and AMIs

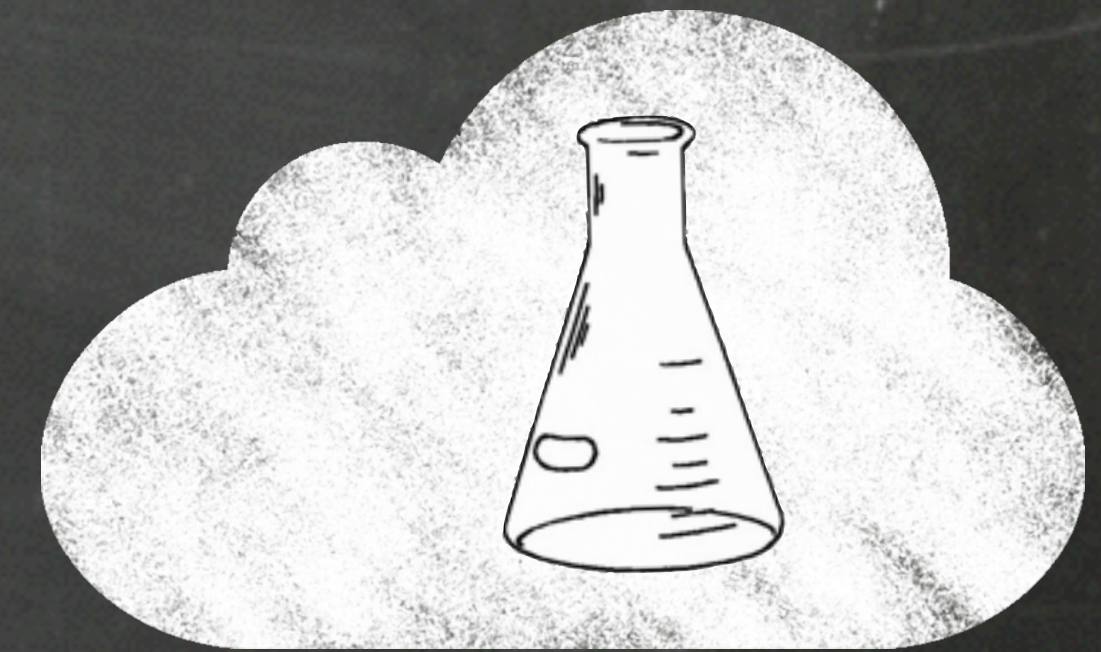


Where is this Going?

1. Researcher conducts experiment
2. Experimental data and results uploaded to the cloud along with reproducible machine images
3. Reviewers leverage cloud resources to reproduce and validate results.
4. Results published in a peer-reviewed journal, including references (e.g. DOIs) to cloud data and AMIs
5. Other researchers use these resources as a jumping off point for further research, also publishing their results in the cloud.

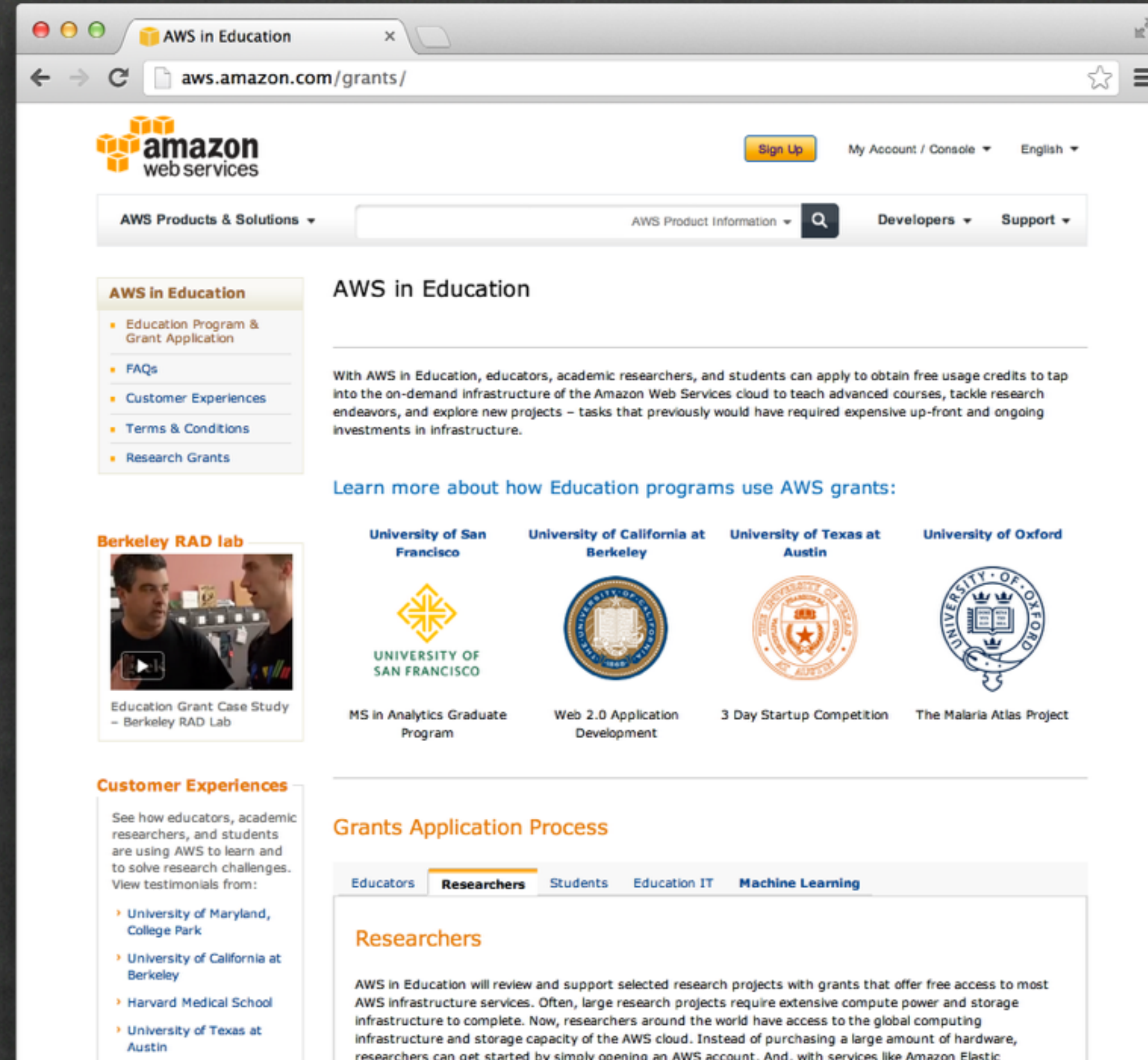


Where is this Going?



1. Researcher conducts experiment
2. Experimental data and results uploaded to the cloud along with reproducible machine images
3. Reviewers leverage cloud resources to reproduce and validate results.
4. Results published in a peer-reviewed journal, including references (e.g. DOIs) to cloud data and AMIs
5. Other researchers use these resources as a jumping off point for further research, also publishing their results in the cloud.
6. Automated workflows re-run the original researcher's experiments on the new data using the original machine images. Interesting results trigger notifications and further review.

AWS Academic Grants



AWS.amazon.com/grants

Thank
You

